

THIS WEEK

EDITORIALS

EBOLA The lethal jungle menace that is once more on the march **p.126**

WORLD VIEW Hidden plagiarism allows cheats to prosper **p.127**



LOCOMOTION Kangaroo tails help them skip through outback **p.128**

Brain fog

The criticism of Europe's Human Brain Project by leading scientists reflects a messy management structure that is in urgent need of clear direction.

Europe's ambitious project to unpick the workings of the human brain faces a crisis less than a year after it was launched with great fanfare at the Swiss Federal Institute of Technology (EPFL) in Lausanne. Some neuroscientists involved in the billion-euro Human Brain Project (HBP) are furious that much of their research into how the brain executes its cognitive functions is to be sidelined as the initiative enters its next phase.

Arguments over the strategy and direction of mega-science projects are nothing new. But the acrimony over this project is particularly unfortunate, given its status as one of two European Union (EU) flagship programmes designed to cross some of the widest interdisciplinary barriers and solve societal problems — such as brain disease. Already, some leading scientists have walked away. If more follow, the project could waste a golden opportunity to understand the brain.

Dissent in the ranks about what the project should encompass and who should decide this has been raging for months. But it peaked in late May, when the project's leaders made clear that they intended to exclude studies on cognition from their core future plans.

The first funding, or 'ramp-up', phase of the brain project began in October last year with €54 million (US\$73 million) from the European Commission and is scheduled to run for three years. The second phase of the ten-year project will be funded to the tune of around €100 million per year for two or three years. But in their detailed plans for this second stage, submitted on 10 June to the commission for approval, the project managers eliminated research on human cognitive architecture.

Cognitive scientists feel ousted. As we report on page 133, they criticize the management of the project — which is hosted at the EPFL and involves 80 or so partner institutes — and say that it has lost its chance to unite EU neuroscientists behind its self-proclaimed 'Moon-shot-like' attempt to understand the brain.

Appropriately, given its subject matter, the workings of the project are complex. The funds come not from the commission's research arm, but from its directorate on information and communication technologies (ICT). As such, the selection of the HBP was on the basis of its scientific merit and the likelihood that computers would be used to solve important problems.

BEST-LAID SCHEMES

The HBP proposed building "a completely new ICT infrastructure for neuroscience" to help to address the problems of brain disorders. But from its inception, the project, led by Henry Markram at the EPFL, has been promoted as primarily a neuroscience rather than a technological effort. Even at its launch last October, its press release opened with the statement: "The world's most ambitious neuroscience project is underway."

Markram, himself a neuroscientist, has been a strong proponent of the part of the project intended to simulate the workings of the entire human brain in a computer. To many outsiders, this promise has become

synonymous with the HBP. Yet the idea is controversial in the scientific community (see *Nature* **482**, 456–458; 2012). Many neuroscientists simply do not think that it is possible to simulate the brain — that is, reproduce its signals in detail — with just the 'bottom-up' data (molecular, cellular, anatomical and the like) that the project will collate.

Stanislas Dehaene, director of the INSERM-CEA Cognitive Neuroimaging Unit in Paris and head of the HBP's cognitive-neuroscience efforts,

"The project's advisory boards are pushing for a sensible resolution."

says that such a simulation, "while not totally useless, will fail to elucidate brain functions and diseases, much like a simulation of every feather on a bird would fail to clarify flight".

Along with other ousted colleagues, Dehaene believes that a top-down, reverse-engineering approach is required, starting with behaviour and high-density recordings of electrical activity in the brains of humans or animals to elucidate how information is encoded and used.

Funding plans for the second phase of the project are more complicated because the European Commission wants to promote its European Research Area. Starting in 2016, the project should receive €50 million per year from the commission for phase two. This must be matched by funding from national sources through a system that is still being worked out, but is likely to involve the pooling of cash from funding agencies in EU member states that wish to participate in the project. The pot of money will finance the winners of competitive calls for 'partnering projects'.

According to the HBP's proposal for this phase, cognitive neuroscience will be able to access the project in this way. But it might be unrealistic to expect the best scientists to apply to be partners without a say in the general scientific direction of the programme. Indeed, in an open letter to the European Commission on 7 July, more than 150 neuroscientists pledged to boycott the partnering projects if their concerns are not resolved. This underscores the importance of ensuring that the arguments are resolved immediately, before rifts get wider and deeper.

The HBP scientific advisory boards have recognized that some changes in management structure are needed. Last week, together with the provost of the EPFL, they put forward some proposals to make the management more representative of the programme's participants. They are helpful, but may not go far enough.

Simply, the brain project needs a consensual management system that allows the full spectrum of leading European neuroscientists to line up behind it — and that is even more important given that the scientific premise of the high-profile brain-simulation pillar remains controversial. It may be too soon to commit to a single path.

The present crisis can be solved. Certainly, the project's advisory boards are pushing for a sensible resolution. The second-phase plan will now be reviewed by a European Commission panel, providing a further opportunity to request changes. The flagship brain project must sail on. There is too much at stake for it to face an unnecessary headwind. ■

Barriers to trust

An outbreak of Ebola highlights the difficulties of implementing public-health measures.

The current outbreak of Ebola virus disease in West Africa, which began last December in Guinea and has since spread to Liberia and Sierra Leone, has already caused 779 cases, including 481 deaths. This makes it much larger than past outbreaks, and it has generated extensive media coverage worldwide.

It is not difficult to explain the macabre interest in Ebola. The virus can kill up to 90% of the people it infects. It causes a horrible death, with initial symptoms of fever and muscle pain followed by vomiting, diarrhoea, kidney and liver damage, and sometimes profuse internal and external bleeding. Despite these tragic effects, a sense of perspective is required. Since the virus was first discovered in Zaire (now the Democratic Republic of the Congo), there have been only a few dozen sporadic outbreaks — most of them small and largely confined to central Africa.

Ebola is a rare disease that, in some 40 years, has resulted in barely 2,000 deaths. Although even one premature death is too many, the toll of Ebola pales beside the millions killed every year in Africa by other infectious diseases. These include HIV/AIDS, malaria and influenza, as well as measles, diarrhoea and pneumococcal pneumonia, which rarely get such attention from the media. Ebola also poses a very low risk to travellers and those outside the affected region, as long as they take basic precautions against infection.

Ebola seems to hold little or no pandemic threat. Unlike an outbreak of a new pandemic flu strain, which is impossible to contain,

Ebola spreads poorly between people. Its transmission requires direct contact with the bodily fluids of those who are infected, such as saliva, faeces or blood. This also means that, even in the absence of a licensed Ebola vaccine, stopping an outbreak in its tracks should, in principle, be straightforward. The disease can be controlled by public-health measures alone, such as surveillance and diagnosis of those infected and their contacts, prompt quarantine of cases and other basic infection control.

“Stopping an Ebola outbreak should, in principle, be straightforward.”

It is not known how the current outbreak began. Fruit bats that serve as a probable reservoir of Ebola are present in the remote forest in southeast Guinea where this outbreak was initially reported, and the disease is thought to have jumped to humans through contact with contaminated bushmeat or infected primates. However, the continuing outbreak, which is now affecting dense urban populations, is mainly being driven by human-to-human spread: traditional burial practices that include contact with corpses, sick people not getting or seeking care, and a lack of basic infection-control measures.

The difficulties in implementing the necessary public-health measures in Guinea, Liberia and Sierra Leone, which are among the poorest countries on the planet, are formidable. Because people in these regions are often poor and many are illiterate, it is not easy to communicate the risk factors for Ebola and the infection-control measures needed — particularly when the disease, and the biohazard suits of health-care workers, can strike terror into the hearts of affected communities. Some workers have reportedly been attacked by locals who believed that they had brought in the disease.

As with social resistance to vaccines in even the richest countries, sometimes diseases continue to cause illness and death — not because of a lack of tools to stop them, but because of the difficulties of conveying public-health measures and overcoming fixed, but misplaced, beliefs. ■

Be concerned

A possible link between neonicotinoid pesticide use and a decline in bird numbers is worrying.

Does anything eat wasps? Despite the popularity of the question, which spawned a series of bestselling books, the answer is less interesting than one might think. Yes, lots of things eat wasps — bigger wasps, for a start. Then there is the bee-eater (don't blame it, we chose the name) and hundreds of other bird species. Perhaps one might instead ask, does anything not eat wasps?

Plenty of these birds eat bees, and other insects too. So it is reasonable to ask a different question: if the numbers of these insects fall, will the birds that feed on them have enough to eat?

This is a complicated question, more difficult to answer than it might seem. So the provocative findings published in a paper on *Nature's* website this week must be taken as a question themselves, rather than a definitive answer. In the paper, researchers suggest that common agricultural pesticides that have long been linked to the decline in some bee species could also be affecting birds (C. A. Hallmann *et al.* *Nature* <http://dx.doi.org/10.1038/nature13531>; 2014).

The suggestion is based on an observed correlation between declines in some farmland bird populations and the use of neonicotinoid pesticides in the Netherlands. The analysis indicates that pesticide use may reduce the amount of prey insects available to birds, causing the observed association, and suggests that neonicotinoids pose an even greater risk to wildlife than previously thought.

Correlation, as every science blogger who reports this story will point out, is not the same as causation. The evidence that the

agricultural chemicals can be blamed for the loss of the birds is circumstantial, being based on the finding that the steepest declines are in regions of the country that also show the highest water concentrations of imidacloprid, the most commonly used neonicotinoid. But the authors also rule out confounding effects from other land-use changes or pre-existing trends in bird declines.

Insects form a large part of the diet of many bird species during the breeding season and are important for raising offspring. Nine of the 15 species investigated exclusively eat insects, and all feed insects to their young, which supports the theory that pesticide use is negatively affecting bird populations by depleting their food sources. Still, although many birds do eat bees, none of those studied would ordinarily eat them in any quantity. So a new question must be addressed: if the pesticides are killing the birds, then how are they doing it?

In an accompanying News & Views article, ecologist Dave Goulson suggests one possibility (D. Goulson *Nature* <http://dx.doi.org/10.1038/nature13642>; 2014). Almost all of the neonicotinoid used ends up in the soil and water as a pollutant. Some could be taken up by non-crop plants, such as hedgerows. Surrounded by tainted food and water, aquatic insects and grasshoppers, beetles and caterpillars could be poisoned — just as bees are — and so not be around to feed the birds.

The picture is incomplete, and thus Goulson's invocation of Rachel Carson's *Silent Spring*, a tale of ecological doom, may seem premature. But his is not a lone voice. An international assessment on the likely effect of neonicotinoids, released by scientists (including Goulson) late last month, warned that the consequences for biodiversity and human food supplies could be severe (see go.nature.com/gzhg94).

And Europe has already imposed a two-year moratorium on the use of three common neonicotinoids on flowering crops.

So, does anything eat wasps? Yes, and we must do more to ensure that that continues. ■

➔ **NATURE.COM**

To comment online,
click on Editorials at:
go.nature.com/xhunjv



Not all plagiarism requires a retraction

Papers that plagiarize only text can still contribute to the literature, but any errors or omissions should be prominently corrected, says Praveen Chaddah.

The ease with which large chunks of text can be digitally scanned and compared with what has previously been published has produced a new breed of academic watchdog. Plagiarism-detection software has opened up scrutiny of scientific publications to non-experts and text that has been copied and pasted without proper attribution is now a common reason for papers being retracted. Hobbyists and political opponents have made a cottage industry out of searching the back catalogues of high-profile individuals for evidence of such misdeeds.

Such plagiarism is unethical and it is a form of misconduct, but scientists are not writers. We value the originality of ideas more than of language. There are worse offences than text plagiarism — such as taking credit for someone else's research ideas and lifting their results. These are harder to detect than copy-and-pasted text, so receive less attention. This should change. To help, academic journals could, for instance, change the ways in which they police and deal with such cases.

To scientists, plagiarism of an idea strikes at the heart of research as a creative enterprise. An idea could be a hypothesis to explain observations, or an experiment designed to test a hypothesis. Such plagiarism is difficult to uncover unless errant authors make fortuitous slip-ups, or it is accompanied by text plagiarism. I have been a victim of both.

In the first, scientists published a paper in a physics journal that failed to give me and my colleagues credit for a method that we devised to test our idea of a magnetic glass state. We called it cooling and heating in unequal field (CHUF), and the plagiarism might have gone unnoticed had the errant authors not also copied our protocol's acronym.

After we complained, the journal published a swift correction from the authors, who apologized and acknowledged that we had published the idea first.

In the second instance, another set of authors duplicated one of our ideas, along with large chunks of text. In this case, the correction addressed the text plagiarism only.

There is a third form of copying: results plagiarism. This is different from fraud, in which the claimed experiments are often not carried out. In results plagiarism, scientists can repeat an experiment and obtain valid data. Such reproduction is, of course, a useful and common feature of science. The deception comes when they fail to mention the original work.

Giving credit to those who did the original work is essential. Credit is a driving force for

humans and quality of original thought is one of the most fundamental ways to judge scientists' work and determine who progresses in their career.

In my opinion, scientists who plagiarize ideas and results are usually established researchers who can assess the validity of published work and probably to have publications of their own accepted and cited.

How should this be dealt with? It is important to appreciate why and how scientists may indulge in the three forms of plagiarism. Copied text in a paper's introduction or concluding paragraph may happen simply because the authors lacked sufficient command over the language (usually English) to express the concept in a different way, or had read previous works that left an indelible mark on their

subconscious. (This is a charitable interpretation and does not apply if the copying is extensive). Such examples of plagiarism can be enough for some journals to retract a paper. But who benefits from such a penalty? If the experiments are performed as described and due credit is given, then the results in the errant paper are robust and benefit the relevant field.

Second-hand text in the methods and results sections could indicate a more severe case of plagiarism — of ideas and data. Computer software cannot determine whether this is the case; that requires the eyes of independent experts.

Authors who have plagiarized ideas or results have crossed a serious ethical line and should be sanctioned by their institutes, as should other perpetrators of misconduct. But need the errant papers be withdrawn, as many journals currently

insist? I argue that this withdraws new and useful results from the scientific record.

Speaking as someone who has been plagiarized, I believe that a correction to the errant paper is sufficient. The wording of the correction must make clear that the offence was plagiarism, not fraud, and include reference to the original work. Most importantly, the correction should be attached to the PDF file of the paper, so that every download also carries the correction. This does happen, but not enough: many journals provide only an online link to the correction. If viewers, including future referees, do not notice that link, then the corruption of the scientific flow of ideas and credit continues. If corrections were to be made prominent, it could dissuade plagiarism in future. ■

Praveen Chaddah is a condensed-matter physicist and former director of the University Grants Commission–Department of Atomic Energy Consortium for Scientific Research in Indore, India.
e-mail: chaddah.praveen@gmail.com

IT IS
IMPORTANT TO
APPRECIATE
WHY SCIENTISTS
MAY INDULGE IN THE
THREE FORMS
OF
PLAGIARISM.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/dlpe6j

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

ASTRONOMY

The exoplanets that were not

Some of the first exoplanets identified as candidates for habitable worlds turn out to be mirages conjured up by magnetism on their host star.

Earlier studies looked at tiny changes in the motion of the star Gliese 581 and concluded that at least five planets must circle it. Of these, two planets, dubbed GJ 581d and GJ 581g, were thought to be at a distance that would allow liquid water to exist on their surface. But a team led by Paul Robertson of Pennsylvania State University in University Park studied emissions from hydrogen in the star's spectrum and discovered magnetic disturbances within Gliese 581.

These magnetic changes, as they rotate around the star, mimic the signal an exoplanet would produce, and probably misled the earlier researchers. *Science* <http://doi.org/th6> (2014)

BIOMECHANICS

Kangaroos' tail-powered walk

When kangaroos move slowly, their muscular tails work as a fifth leg.

Despite their reputation for bouncing, kangaroos spend much of their day walking on all four legs while they munch on grass. To determine the tail's role in this gait, Max

Donelan at the Simon Fraser University in Burnaby, Canada, and his colleagues videotaped five red kangaroos (*Macropus rufus*, pictured) as they ambled across a force-measuring platform. The kangaroos' tails propelled them forward and lifted them, providing an average of around one-quarter of the positive mechanical work involved in walking. Per kilogram of body mass, the animal's tail provided as much

useful work as a single leg for a walking human, the researchers note.

Biol. Lett. <http://doi.org/thh> (2014)

CLIMATE SCIENCE

Fly farther for climate benefits

Aircraft could be diverted hundreds of kilometres to avoid creating artificial clouds, providing climate benefits.

These clouds, called contrails, can form when planes fly through very cold and

humid atmospheric regions, and can contribute to global warming. Emma Irvine and her colleagues at the University of Reading, UK, modelled the maximum extra distance that different-sized aircraft could travel to avoid these contrail-forming regions, without the additional emissions from fuel offsetting the benefits.

In some modelling scenarios, small aircraft could add ten times the length of the avoided contrail to their flight distances, whereas very large jets could add three times their contrail length. Re-routing many flights



MAURICIO ABREU/JAI/CORBIS

COMPLEXITY

Urban social networks remain tight

People living in big cities maintain a close network of social connections, and one that is larger than those in smaller urban areas.

A team led by Markus Schläpfer of the Massachusetts Institute of Technology in Cambridge used a database of billions of telephone calls from Portugal and the United Kingdom to measure social connectivity. They found that personal contacts are tightly clustered in cities of all sizes, suggesting that people form close-knit communities no

matter where they live. But the number of contacts per person in big cities grew more than expected, the authors report. Residents of Lisbon had, over the course of 15 months, on average, 11 reciprocated mobile-phone contacts compared with the residents of the smaller Portuguese city of Lixa, who had six.

The authors suggest that networks of human interaction underlie the systematic acceleration of urban growth.

J. R. Soc. Interface 11, 20130789 (2014)



JÜRGEN & CHRISTINE SOHNS/FLPA

to travel 100 km further could provide climate benefits, if this avoided contrail formation as a result.

Environ. Res. Lett. 9, 064021 (2014)

ZOOLOGY

Wasps build chanel houses

A newly described species of wasp protects its nest with the corpses of dead ants.

Michael Staab of the University of Freiburg in Germany and his colleagues placed artificial nests for solitary wasps in the Zhejiang region of China. Of the 829 nests they later collected, 73 had an outer chamber filled with dead ants, as well as several inner chambers containing a paralysed spider for each wasp larva to feed on. This new species (*Deuterationia ossarium* sp. nov.) seemed less vulnerable to parasites than wasps using similar but ant-less nests. The authors suggest this 'bone-house wasp' uses the ants as a nest-protection strategy. *PLoS ONE* 9, e101592 (2014)

CHEMISTRY

Water-repellent chemical sponges

Researchers in Japan have created sponge-like crystals that soak up gases and liquids, but repel water.

Compounds known as porous coordination polymers can store or separate different molecules from a mixture, and produce catalytic reactions within their pores. But the polymers often break down in water. Susumu Kitagawa and his colleagues at Kyoto University designed small organic molecules, which they combined to make a porous crystal. The material has a nanometre-scale corrugated outer surface, which repels water droplets but draws other liquids through it. A previous strategy to make porous crystals

water-repellent has been to attach bulky molecules after the crystal is made, but this has the undesirable side effect of blocking the pores that gases and non-aqueous liquids are supposed to enter. *Angew. Chem. Int. Edn* <http://doi.org/10.1002/anie.201402151> (2014)

PLANETARY SCIENCE

Hit-and-run origin for Mercury

Mercury may have formed as the result of one or more 'hit-and-run' collisions between the many protoplanets in the early Solar System.

Mercury, the closest planet to the Sun, is unusual because its large metallic core lacks a massive rocky mantle like the ones that cover Earth, Venus and Mars. Erik Asphaug and Andreas Reufer at Arizona State University in Tempe modelled impacts between planetary embryos in the early Solar System. They show that Mercury may be a survivor of one or multiple early Solar System collisions with a larger planet that stripped off the mantle of the smaller body, leaving only the iron-rich core.

Although the chance of the smaller body surviving each such collision is unlikely, the authors show that if the planet did survive, it is statistically probable that it would have suffered multiple hit and runs.

Nature Geosci. <http://dx.doi.org/10.1038/ngeo2189> (2014)

METABOLISM

Obesity without diabetes

A protein that breaks down iron compounds in the blood could be the key to why some people who are obese can maintain healthy metabolisms and do not develop diabetes.

Andrew Pospisilik of the Max Planck Institute of Immunobiology and Epigenetics in Freiburg, Germany, Harald Esterbauer

SOCIAL SELECTION

Popular articles on social media

Republished paper draws fire

Many scientists have denounced the republication of a retracted study by Gilles-Eric Seralini and colleagues that linked genetically modified maize (corn) to tumours in rats.

Kevin Folta, a horticultural scientist at the University of Florida in Gainesville, tweeted "Holy lumpy rats! Republished Seralini #GMO paper had no scientific peer review before published in the new journal." In reply, Mary Mangan, an independent bioinformatics researcher based in Boston, Massachusetts, tweeted "So it was more like exhumed than republished".

Environ. Sci. Eur. 26, 14 (2014)



Based on data from altmetric.com. Altmetric is supported by Macmillan Science and Education, which owns Nature Publishing Group.

➔ **NATURE.COM**
For more on popular papers:
go.nature.com/g5wwpj

at the Medical University of Vienna and their colleagues examined liver and fat tissue from both 'healthy' obese people and obese people who were resistant to insulin. The authors found lower levels of a molecule called haem oxygenase-1 (HO-1) in the healthy individuals. Obese mice engineered to lack the gene responsible for HO-1 production remained sensitive to insulin and had fewer metabolic problems than normal obese mice. HO-1 may be involved in inflammation that leads to metabolic disease, and finding ways to inhibit it could lead to new therapies, the authors suggest. *Cell* 158, 25–40 (2014)

BOTANY

Bellows blow plant pollen

Plants in Ecuador and Costa Rica have evolved a remarkable 'bellows' system for blowing pollen onto feeding birds.

Agnes Dellinger and Jürg Schönenberger at the University of Vienna and their colleagues found that the flowers of the tropical genus *Axinaea* contain stamens



with bulbous appendages. When these are grabbed and squeezed in the beaks of passerine birds, a blast of air from the cavity-filled appendages fires a jet of pollen onto the birds' heads and beaks.

The stamens provide a highly sugar-rich food for the birds, rather than the nectar or pollen reward that is more usually offered by flowers. The researchers did not observe insects foraging on the *Axinaea* flowers (pictured), but multiple bird species were seen feeding on them, suggesting that these stamens are a novel pollinator-activated pollen delivery system. *Curr. Biol.* <http://doi.org/10.1016/j.cub.2014.05.011> (2014)

➔ **NATURE.COM**
For the latest research published by Nature visit:
www.nature.com/latestresearch

SEVEN DAYS

The news in brief

RESEARCH

Facebook furore

Proceedings of the National Academy of Sciences (PNAS) has officially expressed consternation over an experiment by Facebook. In the study, published in *PNAS* last month, Facebook manipulated the posts that its users could see to test whether seeing only happy or only sad posts would affect a user's mood (A. D. I. Kramer *et al. Proc. Natl Acad. Sci. USA* **111**, 8788–8790; 2014). A *PNAS* statement, released on 3 July, criticized the social-media website for failing to receive informed consent from its users before experimenting on them. But the journal conceded that because Facebook is a private company, it is not bound by the same ethics rules as government-funded researchers.

POLICY

Animal research

Government inspections of animal research in the United Kingdom must be reviewed, according to a 2 July report from an independent advisory group of scientists called the Animals in Science Committee. The report was commissioned after allegations of poor practice at one of the country's most prestigious universities, Imperial College London. See go.nature.com/zutrnx for more.

Future Earth hubs

The Future Earth initiative, an effort to coordinate environmental research around the world, will be led by a secretariat that is distributed across five hubs, the initiative announced on 2 July. The hubs, located in Canada, France, Japan,



Reefs disappearing

Most Caribbean coral reefs could disappear in the next 20 years, according to a report by the Global Coral Reef Monitoring Network, the International Union for Conservation of Nature and the United Nations Environment Programme. *Status and Trends of Caribbean Coral Reefs: 1970–2012* attributes the problem to the loss of the region's two main grazers: the parrotfish and the sea urchin. Disease caused a mass sea-urchin die-off in 1983 and over-fishing has greatly reduced the parrotfish population. The Caribbean has about one-sixth of its original coral cover left, says the report, which analysed more than 35,000 surveys conducted at 90 Caribbean locations. Carysfort Reef — within the Florida Keys National Marine Sanctuary — is pictured.

Sweden and the United States, will jointly coordinate and facilitate research activities; communicate and distribute knowledge; and build research skill in poor countries. The secretariat should be up and running by the end of 2014.

Antibiotic review

The United Kingdom will review its regulatory and research practices to combat the growing global threat of

antibiotic resistance, Prime Minister David Cameron said on 2 July. Cameron has commissioned an independent review, led by economist James O'Neill and co-funded by the Wellcome Trust biomedical-research charity in London, that will examine the economic impact of antibiotic resistance and work out a plan for government action. This could include the creation of regulatory incentives

for companies to develop new antibiotics, increased government investment in antibiotic development, and strategies for cooperating with low- and middle-income countries.

Brain-project strife

Around 150 scientists signed a letter delivered to the European Commission on 7 July protesting about the direction of the European Union's €1-billion (US\$1.4-billion) Human Brain Project. The letter requests that the commission seriously consider whether the programme is still fit for purpose. Launched last October, the project has come under fire from neuroscientists who claim that poor management has run part of the effort's scientific plans off course. See pages 125 and 133 for more.

Development goals

The United Nations says that “much more effort is needed” to meet its Millennium Development Goals — internationally agreed targets such as reducing ill-health, environmental degradation and poverty by 2015. An annual report released on 7 July showed good progress in reducing poverty and fighting malaria and HIV/AIDS, but emphasized that progress on fighting hunger has slowed and more work is needed to tackle child mortality.

BUSINESS

Novartis charged

Prosecutors in Tokyo announced last week that they would pursue charges against a Japanese subsidiary of drug giant Novartis and one of the company's former employees, Nobuo Shirahashi. The charges concern allegations of data manipulation related to research on the

BOB STENECK (1975); PHILIP DUNSTAN (2004)

CHIMP HAVEN blood-pressure medication Diovan (valsartan). In a statement, Novartis said that it had “already undertaken decisive action to address problems” with research programmes in Japan, and that it was “committed to changing the culture” at its subsidiary. The firm, which is headquartered in Basel, Switzerland, also said that it would cooperate fully with the Japanese authorities.

Biotech acquisition

Biotechnology company Genentech announced on 1 July that it would pay US\$725 million for Seragon Pharmaceuticals of San Diego, California, which is developing drugs for hormone-dependent cancers. Seragon focuses on therapies that selectively target individual oestrogen receptors and mark them for degradation. Its lead compound is in early clinical trials against advanced breast cancer. Genentech, based in South San Francisco, California, said that the deal is expected to close in the third quarter of this year.

EVENTS

Retired chimps

More than 100 chimpanzees that had been previously used for biomedical research by the US National Institutes



of Health (NIH) have arrived at Chimp Haven, a federally funded sanctuary in Keithville, Louisiana. (One of the chimps, named Sassy, is pictured.) Their retirement is the first step in scaling back the NIH's primate research. In June 2013, the agency announced that it would retire around 310 of its 360 research chimpanzees to sanctuary. The decision followed a December 2011 report by the US Institute of Medicine that recommended strict limits on the use of chimps in biomedical and behavioural research. See go.nature.com/iixvc for more.

Carbon satellite

NASA successfully launched a carbon-tracking satellite on 2 July, five years after its first attempt crashed into the ocean. The Orbiting Carbon Observatory-2 (OCO-2) will track carbon dioxide from space with the best resolution ever achieved

(see *Nature* **510**, 451–452; 2014), supplementing global measurements already being made by Japan's Greenhouse Gases Observing Satellite. The first OCO launch failed in 2009 when the clamshell-shaped cover for the satellite did not separate from the Taurus launch vehicle. OCO-2 was launched aboard a Delta rocket from the Vandenberg military base in southern California.

Ebola emergency

At an emergency meeting in Accra, Ghana, on 3 July, health officials from 12 countries in West Africa agreed that they need better coordination, financing, community participation and surveillance to contain the ongoing large outbreak of Ebola virus disease. By 2 July, there had been at least 779 cases of Ebola virus in the region, including 481 deaths, according to the World Health Organization. See page 126 for more.

FACILITIES

Accelerator funding

The European Synchrotron Radiation Facility won approval on 3 July for a €150-million (US\$204-million) upgrade, to be completed by 2022. A replacement synchrotron (which uses magnets to accelerate electrons and

COMING UP

10 JULY

The United Nations releases its 2014 *World Urbanization Prospects* report, containing updated figures on the global rise of urbanization, including estimates of the most populous cities and projections to 2030. go.nature.com/9omfo1

10 JULY

Indian Prime Minister Narendra Modi will reveal his government's first budget. Some expect him to outline plans to boost economic growth, including increased funding for science and technology.

produce X-rays) will be built in the existing tunnel in Grenoble, France. A separate particle accelerator, the €1.8-billion European Spallation Source, which is scheduled to start producing neutrons by 2019, got the green light for construction in Lund, Sweden, on 4 July. Germany agreed to pay 11% of construction costs — meaning that 97.5% of the project's funding has been secured.

Wind-power loans

The US Department of Energy issued a US\$150-million conditional loan guarantee on 1 July for the offshore wind-power project Cape Wind. If finalized, it would be the first loan guarantee for an offshore wind project in the United States. Slated to begin construction off the coast of Cape Cod, Massachusetts, next year, Cape Wind will be the first commercial-scale offshore wind farm in the United States and will have the capacity of at least 360 megawatts of power.

➔ NATURE.COM

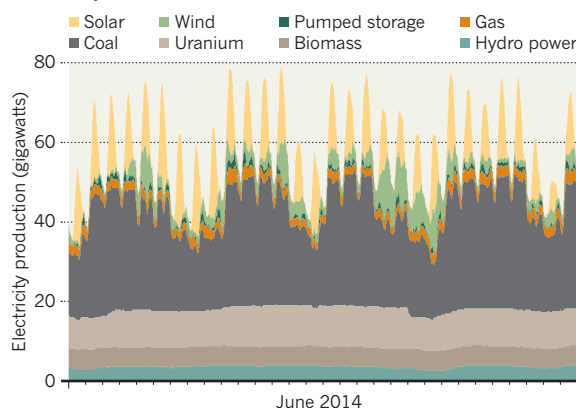
For daily news updates see: www.nature.com/news

TREND WATCH

Solar power produced around 25% of Germany's electrical power at peak times in June, and solar and wind plants together generated 17% of total electricity in the first half of 2014 (up from 13.7% in 2013), according to the Fraunhofer Institute for Solar Energy Systems in Freiburg. Last week, the institute said that all of its electrical-energy data were openly available online (at www.energy-charts.de) so that anyone can track Germany's *Energiewende* — the effort to transform its energy system (see *Nature* **496**, 156–158; 2013).

THE CHANGING FACE OF GERMAN POWER

Online electricity data show high production from solar plants in Germany.



NEWS IN FOCUS

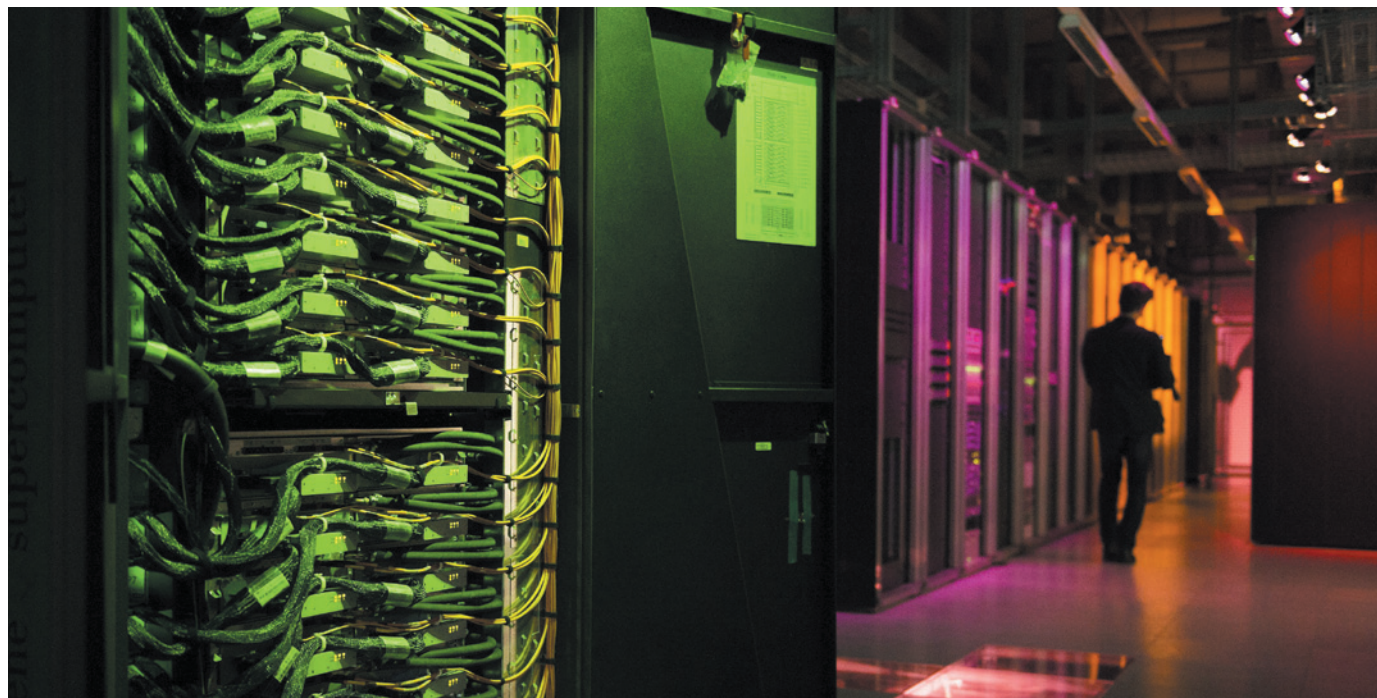
ENVIRONMENT Indian minister's fast approval of development projects worries ecologists **p.134**

VOLCANOLOGY Seismologists plumb depths of Mount St Helens **p.136**

POLICY Drug firms question restrictions on natural-product patents **p.138**

MALARIA CONTROL Dogs hunt for elusive mosquito lairs in Africa **p.144**

JEAN-CHRISTOPHE BOTT/EPH/COREIS



The Blue Gene Q supercomputer in Switzerland is being used in the Human Brain Project to run simulations of rodent brain activity.

NEUROSCIENCE

Row hits flagship brain plan

Changes in scope and focus of European project anger factions of neuroscience community.

BY ALISON ABBOTT

The European Union's high-profile, €1-billion Human Brain Project (HBP), launched last October, has come under fire from neuroscientists, who claim that poor management has run part of the effort's scientific plans off course.

Around 150 scientists have signed a protest letter that was delivered to the European Commission on 7 July. The letter requests that the commission seriously consider whether the project is still fit for purpose as it reviews proposals for the second round of funding, to be awarded in 2016.

The HBP was originally designed to promote digital technologies by supporting and learning from neuroscience. A key element of the project, which has inspired other

brain-research initiatives around the world (see *Nature* **503**, 26–28; 2013), is to develop supercomputers that neuroscientists will use to try to simulate the brain. But as the initiative has developed, its goal has become more and more diffuse. And after months of often fractious discussions about the programme's scientific scope, tempers boiled over at the end of May, when the HBP's three-man executive board decided to cut parts of the project, including one on cognitive neuroscience, from the second phase — in a manner that the signatories say was autocratic and scientifically inappropriate.

Stanislas Dehaene, director of the Cognitive Neuroimaging Unit run by the French Institute of Health and Medical Research (INSERM) and the French Alternative Energies and Atomic Energy Commission (CEA)

in Paris and one of the winners of this year's prestigious Brain Prize, had led this part of the effort. On 30 May, he withdrew his participation from the second phase, citing lack of confidence in some of the decisions being made and in the programme's management; he has not signed the letter.

The escalating row has dismayed the HBP's internal and external advisory boards, which had hoped to resolve tensions that, they acknowledge, arose partly from non-transparent management. Sten Grillner, a systems neuroscientist at the Karolinska Institute in Stockholm and a member of the internal advisory board, says that it is “disappointing” that the issue has exploded so publicly. “I hope it will not be damaging,” he adds.

The HBP is one of the European Union's two Future and Emerging Technologies

► flagship programmes, which are designed to promote information and communication technologies through interdisciplinary research. The project is partnered by around 80 universities and research institutes and its work is organized into three broad interlocking sections: computing, neuroscience and medicine. The cognitive-neuroscience subproject addresses how the brain contributes to tasks such as generating and controlling emotion and making decisions.

The HBP's coordinator, neuroscientist Henry Markram of the Swiss Federal Institute of Technology in Lausanne (EPFL), says that the criticisms represent a minority view of HBP participants and that accusations of lack of transparency are "entirely groundless". "It would be difficult to be more transparent or responsive to our members than we are," he says. He declined to comment specifically on the letter.

Still, he agreed to implement recommendations made by the HBP's advisory boards last week in their bid to diffuse tensions. The boards say that the chair should be elected by the research board — currently the leaders of the 13 scientific subprojects — and should not be a subproject leader, to avoid conflict of interest. They also recommend that the research board elect the executive committee for terms limited to three years.

But some do not think that these measures

are sufficient. Cognitive neuroscientist Zachary Mainen, director of the Champalimaud Neuroscience Programme in Lisbon, who helped to organize the protest letter says that they do not deal with a fundamental failing. The HBP should represent the views of all its members and the neuroscience community at large, he says — not just of the executive board.

The letter, signed by many leading research-institute directors, some of whom are not connected with the project, calls for the review process for the second phase to proceed in an open fashion and for the identity of the reviewers to be made public. It also wants representatives of the reviewing panel on the external steering committee for the period of the funding under review to ensure that the panel's recommendations are put into effect.

BOYCOTT THREAT

The 154 signatories say that if their requests cannot be implemented, the European Commission should reallocate the project's funding — perhaps to the European Research Council, Europe's basic-research funding agency, for broad neuroscience-directed investigator-driven grants. The commission provides only half of the HBP's €100-million (US\$136-million) annual budget; the rest must come from the member states of the European Union through competitive grants. The signatories pledge not to apply for such

funds unless their concerns are addressed.

Preparations for the next round of funding began in January, and the rift between neuroscientists immediately became apparent. Dehaene, for instance, says that he was "dismayed at the unprecedented level of bureaucracy, gobbledegook and absence of transparent democratic reviewing" in the HBP's governance. "There was no need to rewrite the project only months after it came into existence," he says.

The tensions seem to be confined to the neuroscience section of the programme. Physicist Karlheinz Meier of Germany's Heidelberg University, who heads the HBP's computing and robotic section — as well as its futuristic computing platform — says that his section is happy. "I don't see any difference in openness and transparency than in any other mega-project as it approaches a transition stage," he says. "Maybe biologists are less used to projects of this scale than physicists are."

Thomas Skordas, who heads the European Commission's flagships programme, says that the commission closely monitors the progress of the projects and has the power to intervene if it deems it necessary. In a few months, he says, the commission will publish a policy document that will clarify in detail its expectations regarding governance. ■ [SEE EDITORIAL, PAGE 125.](#)

ENVIRONMENT

India races through environmental approvals

But ecologists worry that fast turnaround will do nothing to improve quality of assessments.

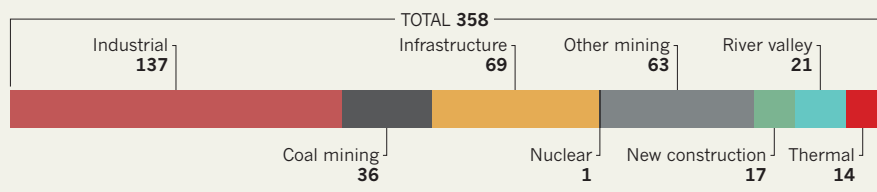
BY T V PADMA

India's new government certainly seems to be cracking on with its job. Within two weeks of taking office, environment minister Prakash Javadekar had rolled out an online system to speed up approvals for development projects that might have environmental impacts. And a system for clearing permits for projects in forests is hot on its heels. But ecologists are concerned that the speed will do nothing to improve an already poor-quality system, and may, in fact, make things worse.

Javadekar had inherited a backlog of hundreds of cases (see 'Projects in waiting') — and pledged to wipe clean the ministry's image as a hurdle to development. But although those

PROJECTS IN WAITING

India's new environment minister is crunching through the backlog of development-project applications, but still has several hundred left to process.



fed up with India's notorious red tape welcome the move, others worry that the environment may suffer.

People applying for development projects such as roads, mines or dams are first told by the ministry what type of environmental-impact

assessment (EIA) they need. (For example, if a planned development is in an area where threatened species live, then a wildlife assessment may be needed.) Then — in a step that pre-dates Javadekar's appointment and that several environmental scientists vehemently oppose



The Narcondam hornbill (*Aceros narcondami*) is found only on Narcondam Island in the Bay of Bengal.

— companies can choose who will do their EIAs.

The ministry says that it will turn around applications that contain the correct, completed EIA in 60 days, and will face a penalty for any delays it does not explain. The previous turnaround was 105 days, but it was not strictly enforced.

Javadekar has also promised that applications that get no objections from the ministry, environmental groups or local residents within two months will automatically be approved. Holding up clearances, he says, “amounts to not making a decision or action”.

The changes have worried ecologists and wildlife experts. “There is no problem setting deadlines,” says Rahul Kaul, senior director of the Wildlife Trust of India near New Delhi. “If the purpose of online clearances is to bring in transparency in the clearance system, it should be welcome — but I am quite sceptical that the whole review process can be accomplished in two months.”

Javadekar’s pledge for speedier clearances coincides with a series of articles in India’s daily newspapers on an Intelligence Bureau report that accused specific non-governmental organizations (NGOs), activists and members of the media of blocking the country’s development, and said that NGO campaigns had contributed to a 2–3% decline in India’s gross domestic product.

Javadekar is prioritizing defence projects and infrastructure applications that involve India’s border with China. He has, for instance,

approved the second phase of a defence project known as ‘Project Seabird’ to build a naval base in Karwar, on India’s west coast, and a project that was rejected by the previous government to build a radar station on Narcondam Island to monitor the strong Chinese presence on the nearby Coco Islands.

Narcondam Island, part of the Andaman and Nicobar chain in the Bay of Bengal, is on the United Nations Educational, Scientific and Cultural Organization’s ‘tentative’ list — places that nations consider worthy of listing as a world heritage site. It is described as “an outstanding example of ongoing ecological and evolutionary processes” and has geomorphic features that represent “major stages of earth’s history”, the listing says.

Narcondam is home to a number of species endemic to the island chain, including the world’s only population of the endangered Narcondam hornbill (*Aceros narcondami*). Only around 350 individuals are thought to be left on the island, Kaul says, so “even small developments will have an impact”.

The government’s action does not address the real problems with the system, which include shoddy EIAs, says Kartik Shanker, a conservationist at the Indian Institute of Science’s Centre for Ecological Sciences in Bangalore. If applicants appoint their own

“Speedy clearances will magnify the flaws of an already flawed system.”

EIA assessors, one can hardly expect them to be neutral, he says. “Speedy clearances will magnify the flaws of an already flawed system. We are all in favour of efficiency, but we do need to make consultative and considered decisions when they could have long-term consequences.”

Prodipto Ghosh, former environment secretary and now a policy expert at the Energy and Resources Institute in New Delhi, says that sub-standard EIAs often need to be rectified, and that causes delays. The 500-page documents are complicated and are often prepared by consultants who have low technical skills and do a poor job of data collection, analysis and interpretation, he says. Data-recording instruments are often inaccurately calibrated and the data themselves are scattered and cannot be accessed at one place, he adds.

Ecologist Madhav Gadgil at Goa University in Taleigão cites a 2013 report on leases for mining projects in India’s biodiversity hotspot, the Western Ghats. The two-year study carried out by the Centre for Environment Education, headquartered in Ahmedabad, looked at 105 mining operations in the state during the previous two years, and found that EIAs, compliance reports and environment management plans were all “highly deficient in information” relating to major environmental factors such as land-use patterns, water resources, biodiversity, air quality and the potential impact of air pollution on the health of local people.

In 2011, the Western Ghats Ecology Expert Panel, which Gadgil headed, recommended that the government undertake “a radical reform of the environmental clearance process by assigning EIAs to a neutral competent body that does not depend on payment by project proponents”, Gadgil says. It also pushed for public hearings and mandatory involvement of biodiversity management committees. The report was, however, bypassed by another review commissioned by the same government (see *Nature* 504, 200; 2013).

Some environment scientists are also worried that the new government will overturn previous decisions to reject projects with environmental impacts. For example, it has said that it would review a ban imposed by the previous government on mining within one kilometre of wildlife sanctuaries in Goa.

India’s environment ministry, Gadgil says, is “running in the wrong direction”. ■



**MORE
ONLINE**

TOP STORY



Pain is better than being alone with our thoughts
go.nature.com/oty55u

MORE NEWS

- Foundations of mathematics put to music go.nature.com/y8nobw
- Fracking linked to surge in Oklahoma earthquakes go.nature.com/tdy5ij
- Cheap fast test to slow spread of tuberculosis go.nature.com/j6obwq

NATURE PODCAST



Stem-cell papers retracted; vaccine for fungus-afflicted frogs; and high-res displays for wearable tech nature.com/nature/podcast



The 1980 eruption of Mount St Helens removed a 300-metre portion of the mountain's summit.

VOLCANOLOGY

Survey probes volcano's depths

Seismic project aims to reveal plumbing of Mount St Helens.

BY ALEXANDRA WITZE

In one of the biggest-ever seismology deployments at an active volcano, researchers are peppering Mount St Helens in Washington state with equipment to study the intricate system of chambers and pipes that fed the most devastating eruption in US history.

This month, they will even set off explosions to generate their own seismic waves. The work could inform research into how volcanoes work throughout the Pacific Northwest and in similar geological settings around the world.

Researchers have already explored magma structures beneath volcanoes such as Italy's Etna and Vesuvius, and an ongoing multidisciplinary

project in the Andes targets two volcanoes in Bolivia and Chile where the ground has been rising for about two decades. But the Mount St Helens study is unique because it aims to produce a three-dimensional picture all the way through and beneath the volcano. If all goes well, it may probe as much as 80 kilometres underground — deep enough for scientists to visualize the mountain's geological origins.

The volcano is part of the Cascade mountain range, which, like other volcanic ranges such as the Andes, rises where a plate of oceanic crust dives deep beneath the lighter-weight continental crust. The diving slab releases water that percolates upwards, lowering the melting point of the rock and allowing magma to form and fuel volcanoes above. Geophysicists would like to understand the entire process better.

"Volcanoes are difficult things to image," says Alan Levander, a geophysicist at Rice University in Houston, Texas, who leads one portion of the project. "If we can get a top-to-bottom image of a connected active magma system, that would go a long way toward understanding."

SECRETS OF DESTRUCTION

Mount St Helens erupted in May 1980, killing 57 people and blanketing much of the western United States in ash. The blast obliterated the top 300 metres of the mountain and sent an avalanche of 2.5 cubic kilometres of debris — the biggest ever recorded — racing down the valley below. Since then, geologists have poked and prodded the mountain with all manner of research instruments, trying to understand what made it go off when it did.

But surprisingly, they have never quite found out how the underground plumbing that feeds magma to the surface works. "It's clearly not in some big spherical chamber, because we would have seen that through previous experiments," says project leader Kenneth Creager, a seismologist at the University of Washington in Seattle. So far, the best picture of what is happening beneath St Helens reaches only 5–8 kilometres down, as reported in a 2009 seismic study (G. P. Waite and S. C. Moran *J. Volc. Geotherm. Res.* **182**, 113–122; 2009).

"We'd really like the tools to look deeper, and recently we've developed those," says University of Washington seismologist John Vidale. The core of the current project, called iMUSH, or Imaging Magma Under St. Helens, combines two seismic techniques: one explores near the surface and one probes deeper.

The shallower study will be setting off the blasts around the mountain. Led by Levander, this 'active-source' seismic project will use the blasts to create seismic waves, which it will record as they travel through the mountain. Changes in the speed of the waves will reveal discontinuities in the mountain's structure, such as chambers of molten magma lurking in the solid rock. The team hopes that the experiment will map down to the boundary between

INTERNETWORK MEDIA/CORBIS

SOURCE: iMUSH

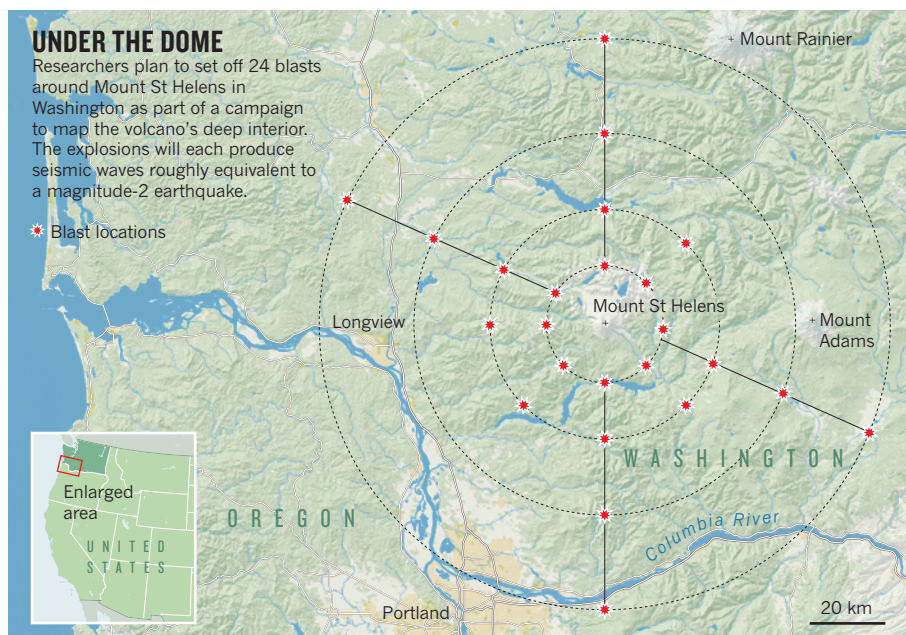
Earth's crust and the underlying mantle, about 40 kilometres deep.

To get the job done, starting next week roughly 65 people will fan out across the mountain to deploy 3,500 small seismometers along roads and back-country trails. They will drill 24 holes some 25 metres deep, drop in industrial explosives used for quarrying, and refill the holes (see 'Under the dome'). The plan is to detonate the explosives in separate shots over four nights. Each blast will shake the ground as much as a magnitude-2 earthquake.

LISTENING DEVICES

Results from the active blasts will be combined with the passive seismic part of the experiment, which is already under way: 70 larger seismometers around the mountain are measuring how long waves from natural earthquakes take to travel through the ground. Their data can be used to probe as far as 80 kilometres down, says Vidale. On 29 June, the experiment captured an earthquake about 24 kilometres beneath the mountain. Quakes deeper than 20 kilometres happen only about once a year, and iMUSH scientists expect this one to provide valuable data on the volcano's underground workings.

A third iMUSH component collects data on magnetic and electrical properties inside Earth, revealing which portions of the volcano's interior are molten or wet. The work may



help to clarify whether Mount St Helens shares a magma chamber with nearby Mount Adams, a controversial proposal that was put forward in 2009 and has yet to be resolved (G. J. Hill *et al. Nature Geosci.* **2**, 785–789; 2009).

Combining all the data may also reveal how much the volcano's magma chamber has

refilled since its most recent bout of activity in 2004–08, a potential indicator of when it could erupt again. “This will be a huge contribution for our field,” says Olivier Bachmann, a volcanologist at the Swiss Federal Institute of Technology in Zurich. “It’s a fantastic chance that we have here.” ■



Drugs based on natural sources, such as sea-snail venom, are becoming harder to patent under US law.

POLICY

Biotech reels over patent ruling

Firms fight for right to patent natural products and processes.

BY ERIKA CHECK HAYDEN

Guidelines that forbid patents on a wide array of natural products, phenomena and principles have many in the biotechnology and pharmaceutical industries worried about the future of their business.

The rules, issued by the US Patent and Trademark Office in March, are open to public comment until 31 July and are reactions to two recent decisions by the US Supreme Court. In March 2012, the court ruled against Prometheus Laboratories in San Diego, California, saying that the company could not patent metabolite levels used to guide drug dosing. Then, in June 2013, the court rejected a patent claim by Myriad Genetics of Salt Lake City, Utah, on DNA sequences linked to breast cancer, opening the door for firms to develop genetic tests of breast-cancer risk. In both cases, the court based its decision on a section of patent code that forbids patenting “laws of nature, natural phenomena and abstract ideas”. The guidance now states that patentable inventions must be “significantly different” from any natural product.

The patent office has decided that because the decisions in these cases built on previous rulings that involved a wide range of natural phenomena and products, the policy should apply to all claims reciting or involving such laws of nature.

But critics say that many existing medicines

would not have qualified for protection under this standard — and that deserving patents are now being rejected. “Not having patents means not having drugs,” says Kevin Noonan, a patent lawyer at McDonnell Boehnen Hulbert & Berghoff in Chicago, Illinois.

Sherry Knowles, a legal consultant in Atlanta, Georgia, and former chief patent counsel at the drug company GlaxoSmithKline, estimates that almost half the drugs approved in the United States from 1981 to 2010 would have been rejected under these guidelines because they were natural, derived from nature, consisted of biological molecules such as antibodies or were vaccines, which often contain natural ingredients such as pieces of genes or proteins.

Knowles and others argue that the Myriad ruling struck down patents only on DNA that has been isolated from nature, and that the patent office’s policy overreaches in covering too many nature-based products and processes. For instance, the guidance instructs patent examiners that a recipe for keeping fruit juice fresh by mixing it with vitamin E might not be patentable because neither the juice nor the preservative have been significantly modified from their natural form. “In the court’s decisions, ‘natural products’ is not limited to DNA,” explains June Cohan, a legal adviser to the patent office.

Patent seekers seem to have been affected by the precedents even before the guidelines

were released. A survey released on 25 June by publisher Bloomberg BNA in Bethesda, Maryland, found that the patent office had denied 40% of the 1,000 or so applications that fall into areas covered by the guidance and filed since April 2011 — most of which were reviewed after the Prometheus ruling. Patents were rejected, for instance, for proteins intended for diagnosis, medicines extracted from marine organisms and a test for detecting certain genetic traits in aquacultured fish.

That worries lawyers such as Hans Sauer, deputy general counsel for intellectual property at the Biotechnology Industry Organization, a trade group in Washington DC. “It’s pretty clear that the Supreme Court intended the Myriad opinion to be a relatively specific and incremental one,” he says. “Is it really the right outcome — that we’re striking down patent applications on industrial enzymes going into laundry detergent, or on antibiotics?”

But others say that the government is acting exactly as it should. “It is doing the right thing for science, medicine and patients,” says Roger Klein, a physician at the Cleveland Clinic in Ohio and chairman of the professional-relations committee for the Association for Molecular Pathology in Bethesda, one of the plaintiffs in the Myriad lawsuit.

The patent office does not have to amend its policy in response to public input, says spokesman Patrick Ross. But it will probably do so, he adds. “We’re already reviewing the comments and looking to see what changes we might make.” There is no set timeline for any changes.

The survey notes that many companies have been able to overcome patent rejections by making it clear to examiners that their invention differs significantly from something found in nature. For instance, the aquaculture application succeeded once it showed how the applicant had used genetic information to breed cod that are more suitable for farming.

“Companies are waiting and trying to see what they could do to avoid a problem,” says Matthew McFarlane, a patent lawyer at Robins, Kaplan, Miller & Ciresi in New York City, who co-authored the survey. “This is the way it always happens when the rules change.” ■

CORRECTIONS

The third reference cited in the News story ‘Risks of flu work underrated’ (*Nature* **511**, 13–14; 2014) was incorrect. It should have read: Herfst, S. *et al.* *Science* **336**, 1534–1541 (2012). In the Editorial ‘Biosafety in the balance’ (*Nature* **510**, 443; 2014), a typographical error meant that the wrong year was given for the 11 September attacks — they happened in 2001. And the News Feature ‘Close collaborators’ (*Nature* **510**, 458–460; 2014) accidentally described IBM’s research site in San Jose, California, as the headquarters.

ALEX KERSTITCH/VISUALS UNLIMITED/CORBIS



KIMIMASA MAYAMA/EPA/ALAMY

CELL-INDUCED STRESS

As a much-hailed breakthrough in stem-cell science unravelled this year, many have been asking: **‘Where were the safeguards?’**

BY DAVID CYRANOSKI

It seemed almost too good to be true — and it was. Two papers^{1,2} that offered a major breakthrough in stem-cell biology were retracted on 2 July, mired in a controversy that has damaged the reputation of several Japanese researchers.

For scientists worldwide it has triggered painful memories of a decade-old scandal. In February 2004, South Korean researcher Woo Suk Hwang announced that he had generated stem-cell lines from cloned human embryos³, creating a potential source of versatile, therapeutic cells that would be genetically matched to any patient. A frenzy of excitement followed this and a subsequent publication⁴, but that didn’t compare with the media firestorm when the results were revealed to be fabricated. The two main cloning papers were retracted⁵, and the careers of some dozen scientists were devastated.

In the soul-searching that followed, ‘research integrity’ became a hot topic, scientists re-evaluated the responsibilities of authorship, and institutions vowed to improve the way that they police their staff. *Nature* and other journals also made promises, saying that they would vet manuscripts more thoroughly. In an Editorial at the time, *Nature* wrote⁶: “Keeping in mind the principle that extraordinary claims require extraordinary proof, *Nature* may in rare cases demand it.”

A year later, when Shoukhrat Mitalipov of the Oregon Health & Science University in Portland claimed to have cloned embryonic-stem-cell lines from monkeys⁷, *Nature* required independent tests to verify that the lines came from the monkey donors. This verification was published alongside the cloning paper⁸. “I applaud what they did,” says Alan Trounson,

the outgoing president of the California Institute for Regenerative Medicine in San Francisco, who helped with the testing.

Then came Japan’s stem-cell case. This January, Haruko Obokata, a young biochemist at the RIKEN Center for Developmental Biology (CDB) in Kobe, Japan, reported in *Nature*^{1,2} that she had converted mouse cells to an embryonic-like state merely by subjecting them to stress, such as physical pressure or exposure to acid (see *Nature* **505**, 596; 2014). The process, labelled stimulus-triggered acquisition of pluripotency (STAP), was so contrary to current thinking that some scientists said they accepted it based only on the reputation of Obokata’s co-authors, who were some of the most trusted names in stem-cell research and cloning.

But the paper¹ that set out the fundamental technique was soon shot full of holes. There was plagiarized text in the article. Figures showed signs of manipulation, and some images

Haruko Obokata tearfully faces the media after she was found guilty of misconduct in April.

were identical or nearly identical to those used later in the same paper and elsewhere to represent different experiments. More damning were genetic analyses that strongly suggested the cells were not what they were purported to be. And although deriving STAP cells was advertised as simple and straightforward, no one has yet been able to repeat the experiment.

Within the space of six months, Obokata was found guilty of misconduct by her institution; well-respected scientists, including RIKEN head Ryoji Noyori, bowed their heads in apology; and both papers were retracted⁹. In the end, the evidence for STAP cells seemed so flimsy that observers began to ask where were the extra precautions and the ‘extraordinary proof’ that had been promised post-Hwang.

The case has reopened difficult questions about the quality of research and peer review, and the responsibilities of co-authors, institutions and journals. It is also making its mark as an example of how not to do things. The episode has already become a “parable in my lab for teaching students about scientific ethics”, says Jeanne Loring, a stem-cell biologist at the Scripps Research Institute in La Jolla, California.

In this article, the news team at *Nature* — which is editorially independent from the journal team that reviewed and published the STAP papers — attempts to find out what went wrong and what can be learned from the case.

CAUGHT IN ISOLATION

The STAP saga has its roots in a contentious hypothesis more than a decade old. In 2001, Charles Vacanti, an anaesthesiologist at the Brigham and Women’s Hospital in Boston, Massachusetts, said that he had found “spore-like cells” in virtually every type of mammalian tissue¹⁰. According to Vacanti, these cells were pluripotent — that is, they could develop into any type of cell in the body — and seemed to lie dormant until activated, perhaps by injury or disease, to regenerate tissue.

Vacanti told *Nature*’s news team in January that by 2006 his laboratory could grow the cells in large numbers, but that they still “were not exceptionally well characterized by us”. That is, the team had not demonstrated pluripotency. This was a job he gave to Obokata, a graduate student who had joined his lab in 2008.

Proving pluripotency is often done by injecting cells into a developing mouse embryo — creating a ‘chimaera’ — and tracking their fate. It is a difficult experiment, and Obokata needed help. “I was looking for the god’s hand of chimaeric-mouse generation,” she said back in January. A Google search led her to famed mouse cloner Teruhiko Wakayama at the CDB, whose lab she entered in 2011 as a visiting professor. After hundreds of failures to get cells derived from adult mice to show up in chimaeras, she and Wakayama switched to

newborn mice as the source of the cells — and the process worked.

By that point, both Vacanti and Obokata were convinced that the stress of the isolation process was creating the pluripotent cells. Obokata said that the idea had come to her while she was taking a bath and reflecting on the stress in her own life.

In the experiments at RIKEN, she used acid to stress spleen cells from newborn mice, and

“THERE HAS TO BE CONTROL, BUT ALSO TRUST IN SCIENCE, OTHERWISE THE SYSTEM BREAKS DOWN.”

she carried out further experiments to characterize their conversion with Yoshiki Sasai and Hitoshi Niwa, two highly regarded stem-cell biologists at the CDB. With the two key characteristics of STAP cells now demonstrated — they were pluripotent and were created using stressful conditions — she had enough data to publish two papers in *Nature* on 30 January^{1,2}.

Obokata became an instant celebrity in Japan, where the media picked up on details such as the ‘Moomin’ cartoon stickers on her lab equipment and the traditional Japanese cooking apron, given to her by her grandmother, that she wore instead of a lab coat.

But within weeks, anonymous observers began noting mistakes in the papers, including evidence of image manipulation, duplications and plagiarism (see go.nature.com/e4dwry). Researchers also started to report that they could not reproduce the supposedly simple experiment.

On 1 April, a RIKEN investigative committee concluded that Obokata had committed scientific misconduct. She maintained that the results were real, but one by one her co-authors withdrew support for the findings. In principle, *Nature* retracts articles only when all co-authors agree, although in rare cases papers can be retracted even if one or more of the authors dissent. In June, Obokata relented and agreed to retract both papers (see go.nature.com/wsfox5). She has not responded to multiple requests for interview since April. She has, however, been invited to participate — under surveillance — in ongoing efforts at RIKEN to verify the original findings.

Should the papers have been published in the first place? Critics have argued that many of the flaws could have been identified beforehand by *Nature* — the easiest, in theory, being a 17-line passage that was taken almost word for word from a 2005 paper¹¹ by another group.

To detect signs of plagiarism, most journals use a service called CrossCheck. It can compare a submitted manuscript with some 40 million published articles from around 100,000 titles, looking for text matches.

Nature editors did use CrossCheck and did not find the match. But the journal from which the text was lifted, *In Vitro Cellular & Developmental Biology — Animal*, had not been indexed by the service at the time. “Although the databases are very large and growing, there are limitations,” explains Rachael Lammey, a product manager at CrossRef in Oxford, UK, which provides the CrossCheck service. Such misses get flagged “a couple times a year”, she says, but there is no way to know how many instances of plagiarism fall through the cracks.

Moreover, identifying the match probably would not have halted publication. Many instances of copied text do not constitute plagiarism and just require citation of the original source. Indeed, the RIKEN investigative committee concluded that the passage — a methodological description — should have cited the original, but that the failure to do so was not misconduct.

The committee was more vexed by instances of manipulated and duplicated images in the STAP papers. Obokata had spliced together gel lanes from different experiments to appear as one. And she had used an image of cells in a teratoma — a tumorous growth that includes multiple types of tissue — that had also appeared in her PhD dissertation. The captions indicated that the image was being used to represent different types of cell in each case. The committee judged that in both instances, although she might not have intended to mislead, she should have been “aware of the danger” and therefore found her guilty of misconduct. Obokata claimed that they were mistakes and has denied wrongdoing.

PICTURE IMPERFECT

Image manipulation and duplication within the same manuscript can be detected, and journals are increasingly checking for such problems. Jana Christopher analyses images in every manuscript before they are accepted by EMBO Press, a journal publisher based in Heidelberg, Germany. She uses a set of automated adjustments created for the image software Photoshop by the US Office of Research Integrity that change attributes such as contrast and colour to make manipulations easier to spot (ori.hhs.gov/actions).

At the behest of the chief editor of *The EMBO Journal*, Bernd Pulverer, Christopher ran tests on the STAP papers without knowing their background. She spotted three problems:

the gel manipulation that was ultimately attributed to misconduct, a seemingly innocent duplicated image mistake and a composite image of cell colonies most probably done to save space.

“The aberrations we saw are fairly typical,” says Pulverer, who reports that around 20% of scanned manuscripts have been found to have such issues since the journal started looking for them in 2011. *The Journal of Cell Biology (JCB)*, published by Rockefeller University Press in New York, has been systematically scanning figures and images in all accepted papers since 2002 and finds about the same rate.

But the journals do not immediately consider a problematic image fraudulent. Spliced gel lanes, for example, are often attempts to present data more clearly and concisely. In most cases, these manipulations are done naively, to create a ‘prettier’, more informative image, says Pulverer. But, he says, “in cases where we don’t obtain plausible explanations and source data for the figure in question, we dig deeper”. At the *JCB*, acceptance is revoked for about 1% of papers, according to the journal’s executive editor, Liz Williams.

Such scanning methods are far from foolproof. Some worry that alerting authors to problems with their images allows would-be fraudsters to improve their forgeries. And although manipulated images might be easy to spot, it is harder to identify duplications, especially when they come from other articles. “Cross-literature comparisons would require very high-powered search algorithms and probably a supercomputer,” says Pulverer. “This has been discussed for a number of years but never moved forward.” In the STAP case, current image-checking procedures would not have caught the problem with the teratoma image — or several other problems with the main paper¹ that surfaced under closer scrutiny.

Philip Campbell, editor-in-chief of *Nature*, says: “We have concluded that we and the referees could not have detected the problems that fatally undermined the papers.” But scientists and publishers say that catching even the less egregious mistakes raises alarm bells that, on further investigation, can lead to more serious problems being discovered.

Many say that the tests should be carried out on all papers. Christopher says that it takes about one-third of her working week to check all accepted manuscripts for the four journals published by EMBO Press. At *Nature* and the *Nature* research journals, papers are subjected to random spot-checking of images during the production process. Alice Henchley, a spokeswoman for *Nature*, says that the journal does not check the images in all papers because of limitations in resources, and that the STAP papers were not checked. But she adds that as one outcome of this episode, editors “have decided to increase the number of checks that we undertake on *Nature*’s papers. The exact

number or proportion of papers that will be checked is still being decided.”

In the face of extraordinary claims, checking images and text is hardly sufficient to test the veracity of results. Independent genetic

“REPUTATION IN SCIENCE IS EVERYTHING. ONCE GONE, IT’S EXTREMELY HARD TO GET BACK.”

analyses^{12,13} proved that the world’s first sheep cloned from adult cells, Dolly, was genetically identical to the mammary-gland cells from which she was cloned. And Snuppy, created by Hwang’s lab, was similarly confirmed to be the first cloned dog^{14,15}. Both verifications took place after publication of the initial results in order to quell debate, but such verification also could be done pre-publication, before a controversy arises.

Nature took that step in 2007, when it solicited an independent test of Mitalipov’s monkey stem-cell lines, which showed them to be true clones. Those tests required coordination between researchers who had to navigate restrictions on sending cells across borders. It “dragged on for months”, says Mitalipov. He subsequently chose another journal, *Cell*, to publish his next big stem-cell cloning paper¹⁶ — demonstrating the process using human cells — partly for that reason. That paper was published 12 days after acceptance. An anonymous critic pointed out mistakes in the final manuscript including image duplication and mislabelled figures, but the authors proved that they were harmless errors (see *Nature* <http://doi.org/mnk; 2013>).

Some researchers say that it would be best if collaborators on a project carried out more stringent verification tests of their own before submitting a paper — something that Wakayama now woefully acknowledges. In the STAP case, however, it is not immediately obvious how this could have been done. Unlike for sheep, dogs or primates, genetically identical mouse strains and matched embryonic-stem (ES)-cell lines are widely available, making it easy to provide samples that look roughly correct. “There is no way to distinguish, genetically, ES cells from STAP cells originating from the same strain,” says Rudolf Jaenisch, a stem-cell biologist at the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts.

That said, post-hoc genetic analyses do seem to be unravelling the STAP riddle. In March, as problems with the papers mounted, Wakayama outsourced genetic sequencing of the purported STAP cells (see *Nature* <http://doi.org/tf8; 2014>). The work focused on a certain characteristic — the spot where a gene encoding a fluorescent protein inserts itself into the genome — to pinpoint the cells’ origin. The results, which Wakayama announced in June, showed them to be different from the mice that had supposedly been used to make them (see *Nature* <http://doi.org/tf9; 2014>).

A step beyond genetic verification would be independent replication of the experiments. That would probably put undue burden on the scientists asked to do the work. But Richard Behringer, a developmental biologist at the MD Anderson Cancer Center in Houston, Texas, says that asking authors directly whether more than one person in the lab has reproduced the results would be one way for the journal “to ensure that all data and images in the manuscript were solid”.

REPLICATION ISSUES

Nature does not disclose communications between editors and authors, but Campbell says that there were four independent groups on the two papers, and “it was our understanding that the work was independently replicated”. When questions first arose about the STAP cells, moreover, co-authors on the papers were adamant that they had overseen replication.

Sasai says that his lab observed the generation of STAP cells. But in fact, he had asked Obokata to replicate only the first part of the STAP process — the expression of a protein, Oct4 — which he recorded with live imaging. At the time, he says, he did not consider “the possibility of a gap between these cells and the derivation of STAP stem cells”.

Wakayama said that he “independently” produced STAP stem cells that looked exactly like ES cells — development beyond what Sasai witnessed — which convinced him that the process was solid. After problems first emerged, he told the *Nature* news team in an e-mail: “I succeeded at RIKEN independently, therefore, I know this result is absolutely true.” Looking back now, though, he realizes that his replication was not completely independent — Obokata was at his side during the entire experiment “and oversaw every step in the process”, he says. Because he moved to a new position at the University of Yamanashi soon after that test, he never characterized the cells and could not rule out the possibility that they had been switched or contaminated. The RIKEN investigative committee found that Sasai and Wakayama, although not involved in the misconduct, carried “heavy responsibility” for what happened.

The co-author who has most confused the issue of replication is Vacanti. Within a week of the STAP papers being published, he sent photos of what he claimed were human STAP



Nobel laureate and head of RIKEN, Ryoji Noyori, was moved to apologize for the mistakes made over the STAP research.

cells to the magazine *New Scientist*. As others failed to replicate the STAP experiment, he told the *Nature* news team in mid-February: “There really shouldn’t be any difficulty. If I can do it, anyone should be able to do so.” In mid-March, he published online a list of tips for making STAP cells “regardless of the cell type being studied”. To date, however, he has produced no additional evidence that he has derived STAP cells in his laboratory. In a statement released on 2 July, Vacanti asserted that although he agreed to the retractions owing to errors in the manuscripts, he is confident that the “core concept” of STAP will be “verified by the RIKEN as well as independently by others”.

For many stem-cell researchers, the most shocking part of the STAP controversy was the involvement of Niwa, Sasai and Wakayama in such troubled work. “Co-authors of a paper like that should have been certain that they can reproduce results independently and in this case they should share responsibility,” says Davor Solter, a developmental and stem-cell biologist at the Institute of Medical Biology in Singapore. Wakayama takes the blame for not making more effort to check Obokata’s work, such as looking at her notebooks, which the investigative committee found to be alarmingly disorganized.

Others sympathize with the researchers, who themselves were duped — whether through negligence or intention — by a junior colleague. “There has to be control, but also trust in science, otherwise the system breaks down completely,” says Maria Leptin, a molecular biologist and director of EMBO. “I cannot watch over every step while they are pipetting. That’s not the point.”

But in addition to lax oversight, Janet Rossant, a stem-cell researcher at the Hospital for Sick Children in Toronto, Canada, and the outgoing head of the International Society for Stem Cell Research, points to “poor reviewing and editing by *Nature*, who were also too ready to publish without verification”. Campbell disagrees. “*Nature* did not let down its guard,” he says.

Some say that the journal should publish reviewers’ comments to clarify the process. Campbell says that the publication of referees’ comments has been considered, but that the disadvantages — which include potential misinterpretations and the desire of many referees to keep their comments confidential — have prevented the journal from embracing this.

“We have to accept that where there is research, there will be research misconduct,” says Paul Taylor, a research-integrity adviser at the University of Melbourne, Australia. Efforts by institutions to train researchers and improve data-management infrastructure might help, “but no policy, no education or training, no administrative requirement, is going to stop misconduct”.

Taylor adds that the focus should be on how an institution responds. In that sense, he says the STAP problem seems to have been a success. RIKEN has acknowledged flaws in its data management and exaggeration in its press release for the STAP papers. Taylor says that its response has been fast, effective and transparent. In the midst of the investigation into Obokata’s work, Noyori instructed all RIKEN labs to review their published work — totaling tens of thousands of papers — for similar types of errors.

Among stem-cell researchers, STAP has become another cautionary tale to add to Hwang’s, with its own set of lessons. For Loring, the story stresses the importance of good record-keeping and the need to enter collaborations with caution. “I lecture my lab members that being an author carries responsibility for the validity of all of the work in the paper. I really try to live by that.” She says that she has removed her name from authorship lists in cases when she could not vouch for the quality of a manuscript.

But for many it is a lesson hard-learned, once again. “Reputation in science is everything,” says Trownson, in a statement that applies no less to journals and institutions than individual scientists. “Once gone, it’s extremely hard to get back.” ■

David Cyranoski is *Nature’s* Asia-Pacific correspondent, based in Shanghai, China.

1. Obokata, H. *et al. Nature* **505**, 641–647 (2014).
2. Obokata, H. *et al. Nature* **505**, 676–680 (2014).
3. Hwang, W. S. *et al. Science* **303**, 1669–1674 (2004).
4. Hwang, W. S. *et al. Science* **308**, 1777–1783 (2005).
5. Kennedy, D. *Science* **311**, 335 (2006).
6. *Nature* **439**, 243 (2006).
7. Byrne, J. A. *et al. Nature* **450**, 497–502 (2007).
8. Cram, D. S., Song, B. & Trownson, A. O. *Nature* **450**, E12–E14 (2007).
9. Obokata, H. *et al. Nature* **511**, 112 (2014).
10. Vacanti, M. P., Roy, A., Cortiella, J., Bonassar, L. & Vacanti, C. A. *J. Cell. Biochem.* **80**, 455–460 (2001).
11. Guo, J. *et al. In Vitro Cell. Dev. Biol. Anim.* **41**, 278–283 (2005).
12. Ashworth, D. *et al. Nature* **394**, 329 (1998).
13. Signer, E. N. *et al. Nature* **394**, 329–330 (1998).
14. Parker, H. G., Kruglyak, L. & Ostrander, E. A. *Nature* **440**, E1–E2 (2006).
15. Seoul National University Investigation Committee, Lee, J. B. & Park, C. *Nature* **440**, E2–E3 (2006).
16. Tachibana, M. *et al. Cell* **153**, 1228–1238 (2013).



Dana hunts for mosquitoes in Mali.

THE GREAT MOSQUITO HUNT

From dogs to balloons, researchers are using unorthodox ways to find out where malaria vectors hide during a long dry season.

BY EMILY SOHN

The armed guards at Mali's Bamako Senou International Airport had never seen a German shepherd before. The only dogs they were familiar with were the small, scrappy mixed breeds that are common in West Africa. So when Dana, a wolf-like purebred from California, stepped off a plane and into the airport in February 2012, eight soldiers surrounded her and her trainer Sapir Weiss, guns raised.

SAPIR WEISS

Weiss, who once trained antiterrorism dogs for the Israeli army, was eager to get Dana outside after 36 hours of bladder-straining international travel that included a seven-hour stopover in Paris. But the soldiers thought the dog's service vest was a suicide bomb. They ordered Weiss to take it off. They demanded to know where Dana's crate was. "Where's the box?" they yelled. "Where's the box?"

To the guards, and most people in Mali, it was inconceivable that a dog could be trained to travel in economy without a crate. Equally improbable was what Dana had come to Mali to do: sniff out mosquitoes to help eradicate malaria.

Dana is part of an ongoing effort to solve a perplexing mystery. Every year, a swathe of the African Sahel region from Senegal to Sudan experiences an extreme dry season that lasts for up to eight months. As surface waters disappear, mosquitoes can no longer reproduce because their

eggs and larvae must remain wet to survive. The number of mosquitoes buzzing around crashes to near zero.

But when the rains come, adult bloodsuckers appear in explosive numbers in as little as three days — a timeline that is hard to square with the fact that it takes at least eight days for these mosquitoes to grow from egg to adult.

The pattern suggests that adult mosquitoes hide somewhere to wait out the dry season, and that possibility points to a tantalizing plan of attack. Every year, malaria makes hundreds of millions of people ill and kills more than half a million, mostly children in Africa. If scientists could figure out where the mosquitoes go when conditions become inhospitable during dry seasons, they might be able to wipe out the insects — and with them the disease they carry — at a point when they are likely to be easy targets.

For decades, the hunt for mosquito hideouts has both enticed and plagued scientists, who have run up against a long list of frustrations. Among those refusing to give up is Tovi Lehmann, a research entomologist at the US National Institute of Health's Laboratory of Malaria and Vector Research in Rockville, Maryland, who, along with dozens of team members in both the United States and Africa, has spent six years and about US\$700,000 trying to find the elusive insects with every method he can think of, including dogs such as Dana.

The potential pay-off is worth the massive effort, he says. "You could imagine visiting villages for less than half a day, targeting those putative sites and basically cutting malaria transmission to the point where it would be trivial."

ELUSIVE QUARRY

It takes 4 hours to drive from the Malian capital, Bamako, to Thierola, an off-the-grid village of about 300 residents and 120 buildings, made mostly from mud bricks and with thatched or mud roofs.

During the wet season, from May or June through to October or November, half a metre of rain falls on the region. Bushes turn green. Millet, maize (corn), peanuts and other crops grow. And mosquitoes arrive. Fast. Lehmann's team has seen mosquito numbers in Thierola surge tenfold within five days of the rain's start¹.

Entomologists have come up with two explanations for how mosquito populations can swell so rapidly before they have had a chance to reproduce. One possibility is long-distance migration on high-elevation winds. Alternatively, the insects might spend the dry season in aestivation, a unique type of dormancy that occurs in some animals that need to survive long dry seasons.

Lehmann's group found an early clue that aestivation might be the answer. At the end of the rainy season in late October 2008, the team anaesthetized almost 7,000 mosquitoes, marked them with poster paint and released them. During collections the next May, they were amazed to find a live adult female with the telltale marks, despite the fact that *Anopheles gambiae* (the complex of species that transmit malaria most efficiently in the Sahel) are known to live for 30 days at most.

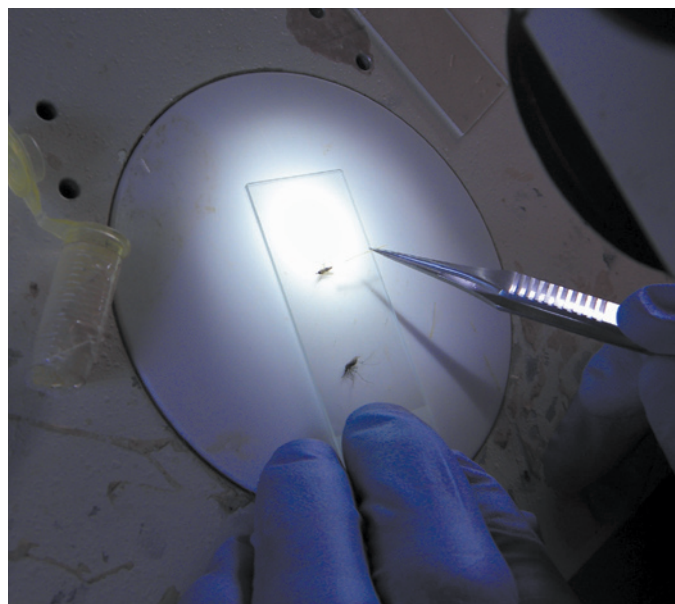
Although aestivation seems a likely scenario, the process has been difficult for biologists to explain. In temperate regions, mosquito species are known to go dormant to survive cold winters, which makes sense because insect metabolisms naturally slow when temperatures drop. Sub-Saharan Africa, on the other hand, is always hot, so it is harder to understand how mosquitoes could slow their metabolism there. The insects must also somehow resist desiccation.

Attempts to induce aestivation in mosquitoes have produced little more than circumstantial and anecdotal evidence. Studies in the 1940s, for instance, tried to replicate natural conditions in the lab but failed to get female mosquitoes to go dormant, according to Douglas Norris, a medical entomologist at the Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland. A letter² published in *Nature* in 1968 described mosquitoes that managed to survive for almost seven months in an insectary in hot, dry Sudan, but those results were never replicated.

Genetic studies could also help to shed light on

➔ NATURE.COM

To watch Dana tracking down mosquitoes, see: go.nature.com/3mcj6



To help dogs to pick up the scent, researchers have been attaching strings doused in vetiver oil to mosquitoes.

the aestivation hypothesis. When Norris and his colleagues looked at genetic markers from one wet season to the next in a village in Mali in the late 1990s, they found that at least 5,000 females had to survive the dry season to found a new population³. And Martin Donnelly at the Liverpool School of Tropical Medicine, UK, and his colleagues have been comparing the genomes of *Anopheles* mosquitoes from across Africa. If genomes remain consistent from one set of rains to the next, that would show even more strongly that many mosquitoes endure the dry season, as opposed to being replaced each year by a population of migrants.

The discovery of the painted female as the rains swept into Thierola showed that a wild mosquito could survive the dry season — the equivalent of a human living for 700 years, Lehmann says. Norris is building field enclosures in southern Gambia for his own studies of mosquito biology and may eventually make try to demonstrate aestivation in that environment. "This is something we believe happens, but nobody has been able to prove it other than Tovi's one mosquito."

So for the next couple of years, as the rains again approached, the team set up nets around suspected refuges, determined to catch mosquitoes as they first emerged. "We thought in a year or two, we were going to be able to find where they hide," Lehmann says. "Everything looked within reach, and it looked very simple."

Despite around-the-clock monitoring, and even a manufactured rainstorm created by a water-filled truck to lure the mosquitoes out of hiding, the insects remained elusive. Potential hiding spots seemed overwhelming, numbering in the hundreds within just 500 metres of the village, making it impossible to put nets and cages around them all. The hunt, it turned out, would not be so simple after all.

ON THE SCENT

Lehmann's project was not the first to misfire. About 15 years ago, Frédéric Simard, a medical entomologist at the Institute of Research for Development, a government institution in Montpellier, France, went on a similar quest in dry-season Senegal. He set up traps around as many potential hiding spots as he could think of, indoors and out, including barns, silos, water-storage containers, wells, tree stumps, tree trunks and cracks in the bottom of dried ponds — to no avail. "There was no evidence for resting mosquitoes anywhere despite huge efforts," he says. "This is basically just like looking for a needle in a haystack."

A group from the University of Miami in Coral Gables, Florida, has scattered tent-like cages around a village in Kenya. And a team based at Hebrew University in Israel is looking for aestivating mosquitoes in



Helium balloons equipped with traps are being used to find any mosquitoes travelling on high-altitude winds.

wet hideouts outside villages in Mali by, among other methods, flooding animal burrows with water to try to flush out the insects. Neither team has found anything conclusive.

Given so many failed attempts before him, Lehmann decided that he would have to come up with a more targeted strategy. When he heard about dogs that could sniff out bedbugs, he latched on to the idea, despite some potentially major challenges. For instance, unlike bedbugs, mosquitoes fly, making them wily and harder to track. And whereas bedbugs emit distinctive pheromones, mosquitoes have no natural scent.

First, he set out to find a way to give mosquitoes an odour. Months of experimenting finally led to vetiver oil, a perfume ingredient derived from an eastern Asian weed grass with an earthy, woody smell. Vetiver is not native to Africa, so it would not confuse a dog's nose. It is not toxic or foul smelling, so researchers would not balk at working with it. And it is stable enough to last for a couple of weeks, long enough to use in experiments. The researchers found that spraying the mosquitoes with vetiver killed the insects, so they soaked tiny bits of string in the scent, then glued the strings onto the bellies of anaesthetized mosquitoes. When they woke, the mosquitoes could still fly, even with the extra weight.

In the meantime, Lehmann called Weiss, who jumped at the opportunity to train a dog to track scented mosquitoes, even though several trainers before him had rejected the idea as impossible. On the sunny lawn of his Olivet Kennel and Dog Training Resort in Santa Rosa, California, Weiss got to work with Dana, allowing her to sniff six cardboard boxes. When she came to the one scented with vetiver, she got her favourite reward: a tennis ball. After a year of training, Dana could find 1-centimetre-long scented strings with 97% accuracy, even when they were put in vials and placed in open holes 20 centimetres deep.

To see whether she would perform as well in Africa, Weiss and Dana flew to Mali in early 2012. Dana sat with Weiss inside the plane so that she would not overheat in the cargo hold. The airport stand-off nearly derailed the whole project, and once they made it to Thierola, Weiss had to retrain Dana because of the extreme heat. Dogs must pant to stay cool, but they can't do that and sniff at the same time. With temperatures near 50°C, Dana needed to learn to stop and cool off every five minutes.

Over the course of a month, Dana found scented strings hidden in

natural holes again and again. She also found mosquitoes that had been marked with scented strings and released days before, including one under a pile of clothes in a laundry basket. The 30 or so Malians in the search crew were flabbergasted. "They were just jumping up and down," Weiss says. "My status got elevated. They thought I was a shaman because I could talk to a dog and the dog would listen to me."

The hope was to get the scented mosquitoes to lead the researchers to their secret lairs. But Dana ran out of time, returning home with Weiss just before a coup swept Mali. Lehmann departed soon after, leaving the mosquito-hunting project in the hands of Adama Dao, an entomologist at the University of Bamako's Malaria Research and Training Centre, who had been leading the Mali-based portion of the team.

In August 2012, two Malian handlers and their dogs travelled to Santa Rosa to learn both the mosquito-sniffing task and how to build a close human-dog relationship. One handler was a physician, the other had a biology degree. Both had been working with Dao and Lehmann for two years on other search strategies, and they were enthusiastic about what they had seen Weiss and Dana do.

The Malian-led dry-season dog work, which continued until December 2013, turned up about nine suspected hiding places, Dao says. Both dogs had independently led the team to the same small tree holes, open holes in the ground near trees and old termite mounds, mostly within about 1.5 kilometres of the village. Nets were set up around the potential shelters at the end of the dry season. But like many who came before them, the researchers found no hordes of aestivating mosquitoes.

A WIDER NET

This month, Lehmann returned to Thierola for the first time in more than two years, armed with some new ideas. He has erected huge nets around the community's wells, for instance. And he has some more-unorthodox plans. Given that other insects might use similar strategies to endure the harsh dry season in the Sahel, he wants to find a larger sentinel species that can be fitted with a radio transmitter and potentially lead researchers to hiding mosquitoes. He also wonders if they could find a marker — such as fungal DNA, trace chemicals or cells from plants or animals — that they could put in suspected shelters. Any mosquitoes emerging from those places would then carry lingering tags that would lead researchers straight to the source, if they can catch the escapees quickly enough.

Meanwhile, the search is moving out and up — to, among other places, a set of caverns about 30 kilometres away and into helium balloons equipped to trap any mosquitoes travelling in by wind. Preliminary studies during the wet season have located a few *A. gambiae* flying as high as 160 metres above the ground, even though there is nothing for them to eat up there. It is now starting to look as though one of the three species of mosquito that transmit malaria in the region aestivates, and the other two migrate, says Lehmann.

So far, Lehmann and his team have published nine papers on the hunt and have three more in the works. As they continue their efforts in the field, other malaria researchers are cheering them on. "Any time you can find a stage where mosquito numbers are very low and they are maybe congregating in one spot, you could really have a big impact," says Gregory Lanzaro, an entomologist at the University of California, Davis, who studies the population genetics of *A. gambiae* in Africa. "There's always hope. When you work on things like malaria, you can't be a pessimistic person. Otherwise, you'd stay home."

Lehmann and Dao remain optimistic that they will eventually find the elusive mosquito dens, where a few low-cost squirts of insecticide could save so many lives. Every setback, they say, provides new information. "We have never been disappointed and we will remain hopeful," says Dao. "As long as we have resources, we will reach our goal." ■

Emily Sohn is a freelance journalist based in Minneapolis.

1. Lehmann, T. *et al.* *Am. J. Trop. Med. Hyg.* **83**, 601–606 (2010).
2. Omer, S. M. & Cloudsley-Thompson, J. L. *Nature* **217**, 879–880 (1968).
3. Taylor, C. *et al.* *Genetics* **157**, 743–750 (2001).

COMMENT



HISTORY Reflections on EMBO, champion of biology in Europe, at 50 **p.150**

SUMMER BOOKS Nature's regular reviewers offer their holiday reads **p.152**

HEALTH Randomized breast-cancer screening trials are ethical **p.155**

METRICS Citations databases need to include more regional journals **p.155**

PHILIPPE LOPEZ/AFP/GETTY



Healthy living needs global governance

Lawrence O. Gostin calls for action on nutrition, pollution and the built environment to curb non-communicable diseases such as diabetes and cancer.

Rapid travel, mass migration and the globalization of culture are known to fuel the spread of infectious diseases. The same factors are also increasing the incidence of non-communicable diseases (NCDs) such as cancer, diabetes and heart disease. As developing countries prosper, these conditions are a by-product of rising air pollution, physical inactivity, and consumption of alcohol, tobacco and excess calories.

More people die from cancer, heart disease and diabetes than from infections, famine and complications of pregnancy in all regions except sub-Saharan Africa. NCDs account for 65% of deaths worldwide and make up more than half the global disease burden¹, a measure of time lost to death and disability (see 'Deadly and neglected'). Some 80% of the 35 million deaths attributed to NCDs each year are in low- and middle-income countries². Globally, the economic cost of unhealthy workforces and surging health-care expenditures is projected³ to be US\$47 trillion by 2030.

Rich foods, cigarettes, fine wines, leisurely lifestyles — and the accompanying health risks — were once the accoutrements of a good life. The poor seemed to have a different set of problems: malnutrition, hard labour and infectious diseases. Yet urbanization and globalized markets are harmonizing cultures and nudging people worldwide towards unhealthy habits. From Dallas to Dhaka, city streets are filled with fast-food restaurants; billboards advertise unhealthy foods and alcohol. Something must be done.

HEALTHY CONVENTIONS

Researchers have identified cost-effective ways to prevent NCDs and the subsequent widening of health disparities. Local and national governments are key to implementing these efforts. Global governance would be a powerful spur — setting norms, mobilizing funding and holding states accountable. This strategy has already worked for preventing deaths from smoking.

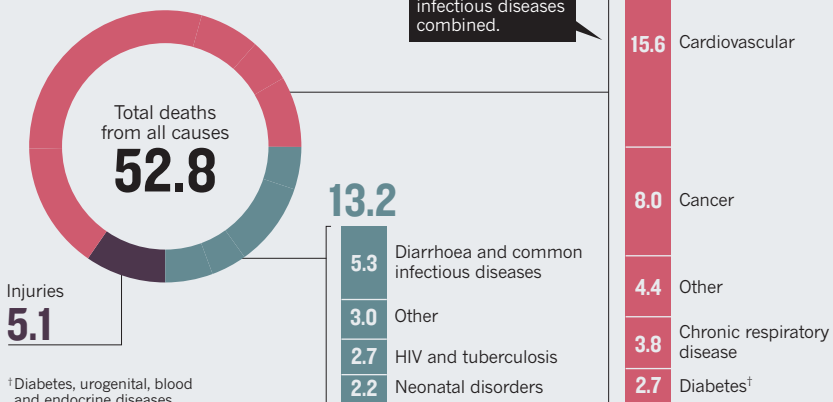
Global agencies should create a dedicated global fund for NCDs, regulate industry to improve nutrition, alter built environments to promote physical activity, and engage diverse sectors of government and other groups in prevention.

The remarkable progress made in curbing tobacco use provides an example of what ▶

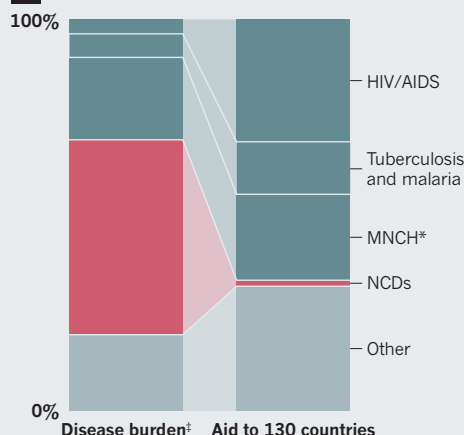
DEADLY AND NEGLECTED

Non-communicable diseases (NCDs) such as cancer and diabetes caused more deaths and disabilities in 2010 than did infectious ones, but received disproportionately low investment.

1 DEATHS (MILLIONS)



2 HEALTH SPENDING



SOURCES: 1, REF. 2, 2, REF. 1

► global regulations can do. Implemented in 2003, the World Health Organization (WHO) Framework Convention on Tobacco Control (FCTC) is one of the most widely ratified treaties in history — 178 nations are now party to convention, accepting requirements to enact robust curbs on tobacco through taxation, prohibitions against advertising, and smoke-free workplaces and public spaces. The WHO offers technical assistance with assessing needs, drafting legislation and exchanging information and technologies.

Although implementation of the FCTC is patchy, it is cited by national courts to justify tobacco-control measures, and the convention has pushed countries towards ground-breaking policies on plain packaging, graphic warnings and sales bans. At least 51 countries have incorporated FCTC provisions into national law, and 79% of parties have strengthened domestic laws governing tobacco⁴.

Tobacco control has saved an estimated 8 million lives in the United States alone in the past 50 years⁵. And the process of creating the FCTC galvanized civil society to hold states accountable, most notably by establishing an alliance to advocate for binding international obligations, and by revealing the extent of the tobacco industry's power⁶.

WEAK AND PATCHY

In every other respect, the global response to NCDs has been weak and fractured. In a systematic analysis of obesity-prevalence research⁷, no country out of the 188 studied had lowered its obesity rates. The WHO devotes only 8% of its budget to NCDs, compared with 39% to infectious diseases and polio eradication. This lopsided distribution of funding is partly because important donors such as the Bill & Melinda

Gates Foundation earmark contributions for diseases such as HIV and malaria that are amenable to concrete targets and easily measured success. These narrow initiatives achieve quick victories, but fail to rein in more widespread health risks.

There has so far been no groundswell of support for NCD prevention. The United Nations did not devote a high-level summit to NCDs until September 2011, by which time it had held three on AIDS. Unlike the AIDS meetings, the NCD summit failed to attract large-scale activism, set targets, take actionable steps or allocate funding. It ignored mental health entirely. The Millennium Development Goals, established by the UN at the start of this century, excluded NCDs; the next set of targets would include NCDs as only a subsidiary goal, if at all.

In May 2012, the WHO did set a global target: a 25% reduction in premature NCD mortality by 2025 (the 25 by 25 goal), followed by a global-monitoring framework



A child at a weight-loss camp in China.

and an action plan. The WHO has also adopted strategies on diet, physical activity, alcohol and mental health (see 'Best buys'). However, these are all voluntary, in stark contrast to WHO treaties on tobacco and infectious diseases.

In 2013, UN secretary-general Ban Ki-moon concluded that national NCD prevention remains "insufficient and highly uneven". This anaemic political response can be attributed, in part, to governments beholden to business interests. The agriculture industry lobbies for subsidies that lower prices of unhealthy foods, for example, for maize (corn) — some of which is turned into high-fructose corn syrup. The food and alcohol industries lobby for low taxation and light regulation — torpedoing New York City's proposed tax on soft drinks, for instance. So too do industries that emit air pollutants. Companies and the media resist advertising limits. And governments find it hard to oversee multinational corporations that span borders.

A PRESCRIPTION FOR ACTION

Without strictures such as those that the FCTC places on tobacco, industries shape the policies that should be reining them in. Food and alcohol companies design and market compelling, unhealthy products, often with misleading labels. Despite peddling large quantities of sodium, sugar and trans-fats, junk-food companies have manoeuvred their way into schools and hospitals. Yet NCDs are often framed as a problem of individual responsibility, with prevention policies criticized as paternalistic.

The damage caused by NCDs goes well beyond individuals. These conditions should be reframed as a collective problem that requires a global response.

CHINA PHOTOS/GETTY

International agreements could establish healthy norms, signalling the priority that countries should give to NCDs. More practically, they could help to mobilize resources, catalyse national policies, set evidence-based standards, and build capacity to draft legislation, share information about best practices, evaluate progress and enforce laws.

The WHO should adopt a convention on NCDs modelled on the FCTC. If politics blocks efforts to establish a broad treaty, the WHO could strategically focus on food and alcohol or adopt 'soft' laws, such as codes of practice with provisions for implementation and reporting. These would go well beyond current guidelines on how much sugar and salt people should eat. There are also models for a hybrid between soft and hard law, such as the WHO's Pandemic Influenza Preparedness Framework, which creates contractual duties on the private sector to contribute vaccines to developing countries.

This week, the United Nations will convene a meeting in New York to assess progress on NCD prevention. Attendees, including representatives from non-governmental organizations, the private sector and the WHO, should consider the following four steps, which would make for robust global governance of NCDs.

Ensure sustainable funding. International funding is needed to build capacity in low-income countries to tackle NCDs with health education, corporate regulation, health-worker training and affordable access to treatment. It could help to assess which NCDs can be attributed to the inability of poor families to buy nutritious foods for young children or expectant mothers.

A global fund for health — modelled on the Global Fund to Fight AIDS, Tuberculosis and Malaria — should be created and commit to a ten-year funding scheme to achieve NCD goals by 2025. This will require major shifts in priorities by richer states and philanthropists.

Regulate industry. Some industry-government collaborations have led to healthier packaged foods. The United Kingdom's salt-intake-reduction programme sets voluntary targets for 85 categories of processed foods. This helped to reduce the population's sodium intake by 15% between 2003 and 2011 and probably contributed to lower rates of cardiovascular disease⁸.

That said, voluntary guidelines often stave off or stall more-effective regulation. In some countries, including the United States, the food industry has adopted weak self-regulatory codes on advertising to children, which has forestalled tougher government regulation.

Such rules are politically difficult to implement. As countries impose higher taxes or stringent controls, they risk losing jobs and

BEST BUYS

The World Health Organization's low-cost, high-impact strategies to prevent non-communicable diseases.

Tobacco	<ul style="list-style-type: none"> • Tax increases • Smoke-free indoor workplaces and public places • Health information and warnings • Bans on tobacco advertising, promotion and sponsorship
Alcohol	<ul style="list-style-type: none"> • Tax increases • Restricted access to retail alcohol • Bans on advertising
Diet and physical activity	<ul style="list-style-type: none"> • Reduced salt content in food • Replacement of trans-fats with polyunsaturated fats • Media campaigns on diet and physical activity
Cardiovascular disease and diabetes	<ul style="list-style-type: none"> • Counselling and medicine for people at high risk of heart attack and stroke
Cancer	<ul style="list-style-type: none"> • Immunization for hepatitis B to prevent liver cancer • Screening and treatment of precancerous lesions to prevent cervical cancer

tax revenues if companies move to more-accommodating jurisdictions. However, if governments acted together, companies could not play regions against each other. Nations should cooperate to create binding agreements that would overcome political disincentives⁹.

Alter the built environment. Healthy activities should be the easy choice where people live, learn, work and play. Bicycle lanes, bike-share programmes, pavements, playgrounds, sports pitches and parks make physical activity safe and attractive. Investing in public transport reduces driving and air pollution. Schools and employers should dedicate time to recreation, offer healthy menus and make stairways more welcoming.

Global governance would facilitate these environments in several ways: by providing templates for redesigning public spaces, schools and workplaces; by offering technical assistance in architecture and city planning; and by creating incentives, such as competitive grants, to promote innovative city landscapes or food programmes. Such governance could promote information exchange, demonstrated by the heart-disease prevention programme in Finland, which included initiatives to increase physical activity and inspired similar prevention measures in other countries.

Prioritize prevention. NCD prevention demands engagement from multiple sectors of government, industry, philanthropy and society. An effective strategy would coordinate ministries of finance, education, labour and urban planning. A good example is the Agita São Paulo ('get São Paulo moving') campaign in Brazil, praised by the WHO and copied in other Latin American countries, encouraging large companies to promote physical activity among their employees¹⁰.

There are multisector examples operating globally: the UN Food and Agriculture Organization harmonizes food and labelling standards and fights for food security through

its Codex Alimentarius standards; and the World Trade Organization adjudicates free-trade agreements. Although tobacco, alcohol and food — risk factors for NCDs — are already under some international control, prevention rarely features prominently in these regimes. Stronger WHO governance would help to counter forces that prioritize short-term economic concerns over NCD prevention and treatment.

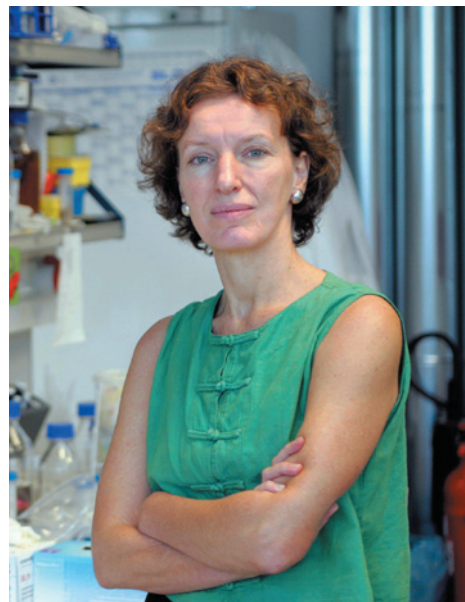
COLLECTIVE RESPONSIBILITY

There are those who argue that the human suffering and economic toll of heart disease, stroke, diabetes, cancer and so on are a matter of personal choice, family responsibility and the free market. But there is a better way to foster healthier behaviour and environments. The NCD crisis is largely of our collective making and can be reversed only through urgent collective action.

What society has to gain from preventing NCDs is not just longevity and prosperity, but also the simple pleasures that come from leading healthier, more vigorous lives — an aspiration well within our grasp. ■

Lawrence O. Gostin is professor of global health law and director of the World Health Organization Collaborating Center on Public Health Law and Human Rights at Georgetown University in Washington DC, USA.
e-mail: gostin@law.georgetown.edu

1. Dieleman, J. L. *et al.* *Health Affs* **33**, 878–886 (2014).
2. Lozano, R. *et al.* *Lancet* **380**, 2095–2128 (2012).
3. Allyn, G. *et al.* *Lancet* **381**, 566–576 (2013).
4. Muggli, M. E. *et al.* *Tob. Control* <http://dx.doi.org/10.1136/tobaccocontrol-2012-050854> (2013).
5. Holford, T. R. *et al.* *J. Am. Med. Assoc.* **311**, 164–171 (2014).
6. Yach, D. *Lancet* **383**, 1771–1779 (2014).
7. Ng, M. *et al.* *Lancet* [http://dx.doi.org/10.1016/S0140-6736\(14\)60460-8](http://dx.doi.org/10.1016/S0140-6736(14)60460-8) (2014).
8. He, F. J., Pombo-Rodriguez, S. & MacGregor, G. A. *BMJ Open* **4**, e004549 (2014).
9. Gostin, L. O. *Global Health Law* (Harvard Univ. Press, 2014).
10. Matsudo, S. M. *et al.* *Pan Am. J. Public Health* **14**, 265–272 (2003).



TOOZE: JOHN TOOZE ARCHIVE; LEPTIN: JUERGEN SCHWARZ/AP/GETTY; GANNON: UDO RINGEISEN/EMBL PHOTOLAB

Directors: John Tooze launched *The EMBO Journal*; Frank Gannon lobbied for better funding; Maria Leptin is forging new alliances.

Fifty years of EMBO

Georgina Ferry reflects on the evolution of the European Molecular Biology Organization, founded to help Europe to compete with the United States.

It began during the Cuban Missile Crisis in 1962. The nuclear physicist Leo Szilard went to Geneva in Switzerland “because he thought America was going to be bombed”, recalls Sydney Brenner, a founding member of European Molecular Biology Organization (EMBO). There, Szilard met Victor Weisskopf, the head of CERN, Europe’s particle-physics lab. “They wanted to found CERB, Centre Européenne de Recherche Biologique,” says Brenner. “Nuclear physics and molecular biology would go together.”

That catalysing moment gave rise to an organization that, taking cues from the Rockefeller Foundation and the Cold Spring Harbor Laboratory in the United States, has acted as a matchmaker, educator, benevolent godparent and advocate for Europe’s life scientists. EMBO’s elected membership has included 79 Nobel prizewinners, and its fellowship schemes have supported thousands of young researchers.

EMBO owes its origin and evolution to the enduring challenge of making European scientists better connected and thence more competitive. How, as it celebrates its half-century, is EMBO remodelling itself for the very different landscape of twenty-first-century life sciences?

In the late 1950s and early 1960s, ambitious molecular biologists were leaving Europe for the United States. In 1958,

Jacques Monod, part of a powerful nucleus of molecular biology at the Pasteur Institute in Paris who went on to win a Nobel prize, warned that the new discipline was forging ahead on the other side of the Atlantic because the structure of European universities put up barriers between disciplines, institutions and countries.

Monod’s proposal for a European institute in Paris went unfunded. In Italy, the geneticist Adriano Buzzati-Traverso was more successful. He established the International Laboratory of Genetics and Biophysics (ILGB) in Naples in 1962, with support from the Italian National Council for Nuclear Research. The ILGB paid higher salaries than Italian universities and attracted researchers from abroad.

Meanwhile, Weisskopf at CERN consulted John Kendrew from the Laboratory of Molecular Biology (LMB) in Cambridge, UK, who had that year received a Nobel prize for his structure of the protein myoglobin. Kendrew immediately saw ‘CERB’ as a way to achieve a level of autonomy that was not available to him in Cambridge. He became its principal advocate and driving force.

In September 1963, European molecular biologists met in Ravello, Italy. A powerful group argued that rather than building a lab, a federal organization should foster interaction by providing fellowships to send scientists to laboratories elsewhere in

Europe, and run regular practical courses where they could learn new techniques such as phage genetics. Buzzati-Traverso supported this proposal, fearing that a second international lab would threaten his ILGB.

Ever the diplomat, Kendrew obtained unanimous votes both to work towards the creation of a lab and to set up a federal organization. The new body would be called the European Molecular Biology Organization. Like an academy, it would elect members on merit. With three years of start-up funds from the Volkswagen Foundation, it was incorporated as a non-profit body in Switzerland on 12 July 1964.

THE RIGHT DIRECTION

The character and influence of EMBO owes a great deal to its directors. The first was the British physicist and radiation biologist Raymond Appleyard. He established and ran EMBO’s fellowship scheme with minimal bureaucracy from his office in Brussels while formally employed by the European Atomic Energy Community (Euratom), a body for the peaceful use of nuclear technology. By the end of the 1960s, 14 countries had come together to fund EMBO’s activities: Austria, Belgium, Denmark, West Germany, France, Greece, Israel, Italy, the Netherlands, Norway, Spain, Sweden, Switzerland and the United Kingdom.

Kendrew finally secured the agreement of

ten of the member states to fund a European Molecular Biology Laboratory (EMBL). With him as its first director, EMBL opened in 1974 in Heidelberg, Germany. It is perhaps not entirely coincidental that EMBL's location on the edge of a pleasant and historic university town bears many similarities to that of the LMB.

EMBL's achievements include the Nobel prize awarded to Christiane Nüsslein-Volhard and Eric Wieschaus in 1995 for their work on early embryonic development. At times, it has been hard for outsiders to grasp the distinction between EMBO and EMBL. What is certain is that neither would have existed without the other.

PUBLISHING AND ASILOMAR

When EMBO, too, moved to Heidelberg in 1973, the British molecular biologist John Tooze took the helm. In 1982, Tooze established and began to edit *The EMBO Journal*, which he ran almost single-handedly until the end of his 20-year term. The journal promoted the EMBO name beyond Europe's borders, and provided a second income stream. It is now ranked nineteenth by impact factor of journals in cell biology and biochemistry.

Tooze took over just as recombinant DNA technology was taking the field by storm. EMBO members, including Ken and Noreen Murray at the University of Edinburgh, UK, ran workshops to introduce European scientists to the new techniques. One of the Murrays' early students was Paul Nurse, who went on to win a Nobel prize and is the current president of the Royal Society in London. "We got lots of hands-on experience and also exposure to some of the great molecular geneticists of the time," he wrote in 2004, in *EMBO: 40 Years of Success*.

In February 1975, after US scientists raised fears about the possible dangers of the DNA technology, a conference in Asilomar, California, agreed a voluntary moratorium on recombinant DNA research. Tooze told the US National Institutes of Health (NIH) that EMBO would be unable to recommend that European researchers adopt the highly restrictive draft guidelines then under consideration, and the organization set up its own recombinant DNA committee. With Ken Murray's help, Tooze organized an experiment to prove that viral DNA was much safer integrated into a bacterial plasmid than it was as part of an intact virus particle (M. Fried *et al. Nature* **279**, 811–816; 1979).

As a result, the NIH held a workshop with EMBO in the United Kingdom, and

subsequently withdrew its draft guidelines that would have required all recombinant research to be carried out in biosafety-level-3 containment. "I think that was a turning point in the regulation of recombinant DNA research in terms of its potential as a biohazard," says Tooze.

POLICY PLAYER

Frank Gannon, a molecular biologist from University College, Galway, in Ireland, took a different approach when he took over in 1994. "I saw EMBO as a way of permeating science throughout Europe with excellence, and of influencing the European Union who were becoming very strong at this stage," says Gannon. To weld EMBO members and fellows into a community, he introduced annual workshops and launched an awards and mentoring scheme called the Young Investigator Programme.

By 2000, the number of countries investing in EMBO had more than doubled and included several Eastern European nations where science was poorly resourced. Many bright young fellows from those countries were making their careers overseas — worsening the state of science in their home nations. So EMBO set up 'installation grants', with support from host countries, to enable returning fellows or researchers to start their own labs. The first countries to volunteer for the scheme were Croatia, the Czech Republic, Estonia, Hungary, Poland, Portugal and Turkey. These schemes, *The EMBO Journal*, two new journals and two big policy programmes to encourage engagement with society and international collaboration saw EMBO grow from 4 to 40 members of staff under Gannon. In 2001, it opened its own building, on land donated by EMBL in Heidelberg.

Next, EMBO went into battle with the European Commission over its policy on grant-making for scientific research. The Framework Programmes of the European Union were 'top-down' funding mechanisms geared towards economic impact. With no European money for bright ideas by individual scientists, there was a growing demand in the scientific community for a European Research Council (ERC), modelled on the US National Science Foundation and any number of national research councils.

BIGGER TENT

EMBO took a lead in lobbying for this change, and the ERC was founded in 2007. In its first five years it disbursed more than €4 billion (US\$5.4 billion) to 2,500 researchers in 480 European institutions. Inevitably a few institutions in a few countries have received a disproportionate share. EMBO and the ERC cling to the principle that all awards should be based on merit alone, and as a result contend with a chorus of complaints

from the countries that feel snubbed.

EMBO's current director is the indefatigable Maria Leptin, a professor at the Institute of Genetics at the University of Cologne in Germany, head of an EMBL lab and president of the lobby group Initiative for Science in Europe. Over the years, career administrators at the tiller of the organization have given way to working scientists who understand the community. They leave the day-to-day running to a professional secretariat set up by Leptin's predecessor, the German molecular biologist Hermann Bujard at the Centre for Molecular Biology at the University of Heidelberg.

The most difficult question that Leptin faces, perhaps more than for any of her predecessors, is what is EMBO for? Its original *raison d'être*, to catch up with the United States in the techniques of molecular biology and to integrate Europe's community in the field, has long since been achieved. Molecular biology has entered the mainstream: few branches of biology can now progress without occasionally manipulating some DNA or solving a protein structure.

EMBO has extended its 1,500-strong membership to new areas such as neurobiology and ecology. Associate members can be of any nationality, and any scientist can apply for an EMBO fellowship to come to a European institution. Countries outside Europe, including South Africa, Taiwan and Singapore, have gained access to EMBO's programmes through cooperation agreements. No longer exclusively European nor exclusively molecular, in 2012 EMBO stopped spelling out its name and adopted the brand 'EMBO: excellence in life sciences'.

EMBO still has problems to solve, old and new. Because of language barriers, pension structures and a host of other factors, it is more difficult for European scientists to move between countries than it is for US scientists to progress around the large number of excellent institutions in their home country. EMBO also has to adapt to shifting career structures. For example, postdoctoral training has changed from a two-year stint to a five-year preparation for independence. EMBO is the first funding organization to have introduced a 'portable pension' for its fellows, and it supports the European Commission's slow progress towards a European Research Area.

The scientific environment in Europe has changed out of all recognition since EMBO's founding 50 years ago. The geopolitical landscape still leaves the organization some mountains to climb. ■

Georgina Ferry is a science writer based in Oxford, UK. She is the author of *EMBO in perspective: A half century in the life sciences*.
e-mail: mgf@georginaferry.com



SUMMER BOOKS

As the wild blue yonder beckons and labs and classrooms empty, *Nature's* regular reviewers share their holiday reads.

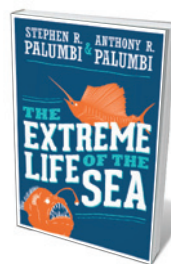
EDITORS' CHOICE

1

THE AUTISTIC BRAIN / TEMPLE GRANDIN & RICHARD PANEK

2

WHITE BEECH / GERMAINE GREER



The Extreme Life of the Sea

STEPHEN R. PALUMBI AND
ANTHONY R. PALUMBI

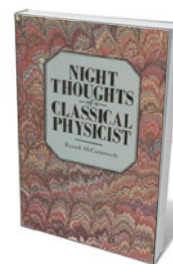
Princeton University Press: 2014.

As generations of inspiration-seeking fiction writers can affirm, nature's strangeness stretches the limits of our imagination. This gem of a book by marine biologist Stephen Palumbi and his son, science writer Anthony Palumbi, finds enough weirdness in the ocean to feed creativity for generations to come.

One of my favourites is *Osedax mucofloris*, the "bone-eating snot flower" — a microbe-filled worm with no mouth or anus that lives on the bones of dead whales on the deep seabed. It looks, they write, "like nothing so much as the contents of a tissue after a sneeze".

The Palumbis' writing is a wonderful mix of meticulous science and creative panache. So, albatrosses have "vast white wings and the inner peace of a seraph", can fly 800 kilometres a day over open ocean, surfing "on the thin current of air pushed ahead by the wave". Meanwhile, the lumpy knobs of flesh that disrupt the smooth leading edge of a humpback-whale flipper reduce drag — a trick that seems to work for wind-turbines too. Coming soon to a hillside near you: turbine blades like whale flippers. A joy whether read at one sitting, or dipped in and out of to prolong the pleasure.

Callum Roberts is a marine conservation biologist at the Environment Department, University of York, UK.



Night Thoughts of a Classical Physicist

RUSSELL MCCORMMACH

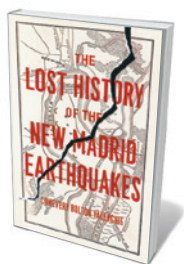
Harvard University Press: 1982.

Germany, 1918: a classical physicist is having night thoughts of loss. His starving country is losing the Great War; his institute is rife with infighting; he easily loses focus and has fainted during a lecture. Worst of all, he thinks, his clean, logical, orderly physics is dissolving into chaos.

Russell McCormmach is a historian of physics, and this book is its own genre: a novel constructed from scholarship, whose main character is a composite of historical physicists and in which every conversation, letter, room and thought is based on historical evidence.

The classical physicist worked with the mechanical world view, where light travels in waves through the ether's bright fluid and effects have causes. The modern physics of Max Planck and Albert Einstein said that light is both wave and particle, and no ether exists. Later, Werner Heisenberg said that the world at its most fundamental is uncertain. This new physics must have been disorienting, demoralizing, even frightening. The novel's great virtue is how thoroughly the reader understands the anxiety that, as Planck wrote, is inherent in resisting meaninglessness and replacing it with mere incomprehensibility.

Ann Finkbeiner is a freelance science writer in Baltimore, Maryland, and author of *The Jansons* and *A Grand and Bold Thing*.



The Lost History of the New Madrid Earthquakes

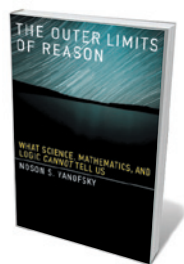
CONEVERY BOLTON VALENCIUS
University of Chicago Press: 2013.

Crack open this book and you are racing through Tennessee woods, tracking a bear with frontiersman Davy Crockett on a cold night in 1826. When the animal tumbles into one of the region's "earthquake cracks", Crockett dives after it, slaying the wedged bear with a knife.

Geology matters. As historian Conevery Bolton Valencius reveals, the trap was an effect of the unusual intraplate earthquakes of 1811–12, epicentred near New Madrid in Missouri's southern 'Bootheel'. Powerful enough to break glassware in Baltimore, Maryland, the quakes transformed the landscape of the middle Mississippi River, turning cropland into swamp.

Big chunks of US history have fractured along the zone's fault lines. The Bootheel was where the agrarian South of the eighteenth century abutted the industrial nineteenth-century North; the muddy Mississippi divided settled East from wild West. Weaving deep time with human time, Valencius gives us exemplary science history: accurate yet erudite, entertaining but substantial, adroitly marshalling the past to interpret the present.

Nathaniel Comfort is professor of the history of medicine at Johns Hopkins University in Baltimore, Maryland. His most recent book is *The Science of Human Perfection*.



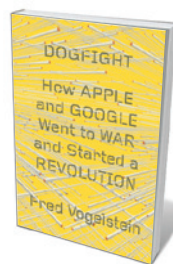
The Outer Limits of Reason: What Science, Mathematics, and Logic Cannot Tell Us

NOSON S. YANOFSKY
The MIT Press: 2013.

"The last function of reason," wrote French mathematician Blaise Pascal in the posthumously published *Pensées* (1669), "is to recognize that there is an infinity of things that are beyond it." That domain of ignorance is the subject of this unusual popular science book by computer scientist Noson Yanofsky. He follows in the footsteps of UK Astronomer Royal Martin Rees, who spoke eloquently on the topic in his 2010 Reith Lectures.

Steering well clear of complex mathematics, Yanofsky canters through language paradoxes, infinity puzzles, computing impossibilities, the scope of fundamental physics and what he believes to be the philosophical limitations of science. It is enjoyable ride, although I suspect that the author underestimates humanity's intellectual ingenuity. Pascal would, I believe, have been astonished by the extent of scientific progress over the past 350 years, especially now that subjects once out of reach, such as early-Universe cosmology, have gradually moved to within the gaze of enquiry and reason.

Graham Farmelo is a by-fellow at Churchill College, University of Cambridge, UK, and author of *Churchill's Bomb*.



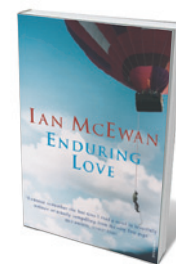
Dogfight: How Apple and Google Went to War and Started a Revolution

FRED VOGELSTEIN
Sarah Crichton: 2013.

Today, two mobile operating systems — Apple's iOS and Google's Android — run on most mobile phones and tablets, and are moving to other smart devices. These corporate superpowers battle directly for people's minds and pockets through proprietary ecosystems — devices, applications, services and content. Yet until recently, Google and Apple were business buddies: Apple's late co-founder Steve Jobs was mentor to Google counterparts Larry Page and Sergey Brin, for instance. What happened?

In August 2005, Google, intent on building an open mobile operating system for handsets, bought a start-up: Android. According to Walter Isaacson's biography of Jobs, Apple had begun work around the same time on the iPhone, because Jobs worried that handsets with good music players would outcompete the iPod. Thus the two companies inadvertently started on a collision course. In *Dogfight*, Fred Vogelstein deploys interviews with executives and key engineers from both companies to tell a refreshing inside story. If anyone wants to see why Silicon Valley still dominates global innovation, start here.

Li Gong is chief operating officer of the Mozilla Corporation.



Enduring Love

IAN MCEWAN
Jonathan Cape: 1997.

In the celebrated opening scene of Ian McEwan's *Enduring Love*, science writer Joe Rose is part of a group of men struggling at the ropes of a hot-air balloon — with a child in the basket — to keep the craft tethered to the ground. When a gush of wind lifts the balloon, they must decide in a split second whether to hold on or let go — and one of the men, the unemployed Jed Parry, is struck by the idea that Joe has fallen in love with him. As the novel unfolds, it becomes clear that Parry's behaviour is a sign of sudden-onset de Clérambault's syndrome, which centres on this delusion.

This slim thriller testifies articulately to McEwan's fascination with neuroscience, as does his novel *Saturday* (Jonathan Cape, 2005), whose protagonist has Huntington's disease.

Enduring Love is a harrowing exploration of the minds of two men, both of them victims: Parry, whose psychopathology propels him on a course of relentless stalking, and Rose, who sees his life fall apart as a result. Most scientific articles on de Clérambault's syndrome fail in helping us to understand the 'inside' of this delusion and the havoc it wreaks in the lives of sufferer and object alike. McEwan's fiction does, in an utterly convincing way. Don't read it late at night.

Douwe Draaisma is professor of the history of psychology at the University of Groningen in the Netherlands, and author of *The Nostalgia Factory*.



The War that Ended Peace: How Europe Abandoned Peace for the First World War

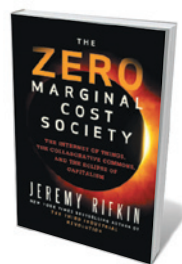
MARGARET MACMILLAN
Profile: 2013.

What mishaps, misunderstandings and mistakes directed so many countries into the First World War? These nations had been marching towards happiness and prosperity through scientific endeavour in communication, transport and energy.

Historian Margaret Macmillan has made this century-old story suspenseful: even knowing the end, we race through her masterful assemblage of people and powers. We visit irresponsible personalities (Kaiser Wilhelm II, all tantrums and ill-informed interference, and the daydreaming UK foreign secretary Edward Grey), unthinking competition (the race to build the most warships), and creaking empires, the Austro-Hungarian and Ottoman hoping, with Russia, that 'a little war' could re-establish their pre-eminence.

The saddest chapter concerns children in uniform and societies convinced that 'every generation needs a good cleansing war'. Armies planned only for a short, glorious attack; they were unequipped for years in the trenches, and a generation was lost to mud, misery and madness.

Margaret Catley-Carlson serves in the Canadian Water Network and was establishing chair of the World Economic Forum's Global Agenda Council on Water Security.



The Zero Marginal Cost Society: The Internet of Things, the Collaborative Commons, and the Eclipse of Capitalism

JEREMY RIFKIN
Palgrave Macmillan: 2014.

Can big data make us happy? Social thinker Jeremy Rifkin thinks so — through the digitized global network the Internet of Things (IoT). In *The Zero Marginal Cost Society*, he posits how people, natural resources and more "will be linked via sensors and software to the IoT platform, continually feeding Big Data to every node — businesses, homes, vehicles". Rifkin sees the IoT as allowing economic exchange through a collaborative commons that will eclipse the capitalist marketplace and save the planet.

Rifkin is known for his insight on social and technological trends. Although he does not sufficiently explain how the IoT will reduce environmental problems, the book has many stimulating ideas on everything from bitcoin and prosumerism to the smart economy. If you are not a millennial and have yet to try crowdfunding, this is for you. Pull up a garden chair — just remember to turn on the Wi-Fi.

Gail Whiteman is professor of sustainability, management and climate change at the Rotterdam School of Management, Erasmus University, the Netherlands.



REVIEWS: GO.NATURE.COM/G7YFL7

The Autistic Brain: Thinking Across the Spectrum

TEMPLE GRANDIN & RICHARD PANEK

Houghton Mifflin Harcourt: 2013.

Temple Grandin, autism's straight-talkin', cattle-handlin' answer to neurologist Oliver Sacks, is back. This time she's playing the lab rat. Grandin takes every new test she fancies to see how science is squaring up to autism — the social, sensory and physical disorder that has more than 1 in 100 people in its grip. Brain scans reveal Grandin's visual cortex to be underwhelmed by faces; an auditory trial confirms that she cannot pick out one voice in a clot of others; even her famed picture thinking is stumped by some spatial tasks. She finds lots of promising leads from neuroscience and genetics, but too few translating to the clinic. Although knowledge about the condition has grown in some ways since the 1940s, "in other ways", she chides, "we're just as confused as ever".

In this, Grandin's eighth book on autism, her prose has been wrangled by the physics writer Richard Panek. The result is better organized than some of her previous works but I missed her unique voice. Plus there is nothing about the malfunctions in digestion and metabolism that some believe are central to autism. Nonetheless, Grandin's tour of what scientists know about autism (embarrassingly little) and how that reflects or alters the lived experience of people on the spectrum and their carers (not enough) is a much-needed call to action.

Sara Abdulla is Nature's chief commissioning editor.

White Beech: The Rainforest Years

GERMAINE GREER

Bloomsbury: 2014.

In 2001, Australian feminist writer Germaine Greer took on a new challenge. "As a lamb to the slaughter", she acquired 60 hectares of Queensland pasture and rainforest. Logged and farmed for a century, the land was now neglected and choked with lantana canes. In *White Beech*, Greer documents how she brought the bedraggled patch back to health and established it as a nature reserve.

Helped by her sister Jane, a botanist, Greer discovers that her abandoned holding is a paradise, filled with nut-festooned macadamia, flowering black bean and magnificent native white beech trees. Once prized for their soft timber, only about 100 white beeches remain in their original habitat. Other inhabitants include platypuses and quolls, pythons, kingfishers and owls. Greer pieces together the land's story, from the Aboriginal tribes who have long lived in the area to the European pioneers who exploited its bounty — from wood to koala skins. Her vivid, often funny, account doubles as a scholarly record of the precious ecology at the Cave Creek reserve.

Joanne Baker is senior comment editor for Nature.

Laboratory Life: The Social Construction of Scientific Facts

BRUNO LATOUR AND STEVE WOOLGAR

Sage: 1979.

It is 35 years since sociologists Bruno Latour and Steve Woolgar unleashed *Laboratory Life* on the world. The book draws on Latour's mid-1970s stint in the lab of brain-hormone researcher Roger Guillemin at the Salk Institute for Biological Studies in La Jolla, California.

Latour engaged in "a research procedure analogous with that of an intrepid explorer of the Ivory Coast". His encampment in Guillemin's soon-to-be-Nobel-prizewinning lab, following bemused researchers at their — to him — arcane daily grind, was designed to capture "the construction of facts". He concluded that this happens through scientists producing papers, then using their status to convince others of the facts' correctness. Thus they gain "credit" that can be re-invested in producing more facts.

Many have objected to Latour's idea that researchers 'create' facts through selective work, rather than 'discovering' truths about the world. (Latour would later cross swords with mathematical physicist Alan Sokal, who famously critiqued cultural studies by publishing a hoax paper in the journal *Social Text*.) But although it is no easy read, *Laboratory Life* remains among the best outsider attempts to study science as a process. The questions arising from Latour's observations — Why are research papers valued? How is their content created? Who is deemed believable in science? — are as relevant today as they were in 1979.

Daniel Cressey is a reporter in Nature's London office.

Candy: A Century of Panic and Pleasure

SAMIRA KAWASH

Faber & Faber: 2013.

Once a luxury, sugar has become ingrained (often as a confusingly named additive) in the modern diet — even as it is increasingly viewed as problematic for our health. In *Candy*, Samira Kawash unwraps the spiralling 'candification' of the United States, providing cultural insights into its twists and turns. We see the nineteenth-century explosion in manufacturing sweets, and early twentieth-century chemistry labs concocting novel products such as Numoline, made by splitting sugar into a mix of glucose and fructose. In the 1920s, 'lunch bars' were marketed as alternatives to a midday meal. One, Sperry Candy's 'Chicken Dinner' consisted of peanuts, corn syrup and milk. We are still hoodwinked by 'healthy' processed foods laden with corn syrup.

Throughout this sweet little history, one theme stands out. The status of confectionary is heavily influenced by the conflicting messages in multibillion-dollar marketing campaigns and the scientific evidence on the risks of overindulging in sugar. I was drawn to *Candy* while shunning sweetened food for six weeks. This book contextualized that challenge and brought home the need for a global look at the uses and abuses of sugar today.

Emily Banham is Books and Arts editorial assistant at Nature.

Off the Map: Lost Spaces, Invisible Cities, Forgotten Islands, Feral Places and What They Tell Us About the World

ALASTAIR BONNETT
Aurum: 2014.

Social geographer Alastair Bonnett is a documentarist of absence, ambiguity and loss. In his wonderful *Off the Map*, he explores 47 places that have evaded the cartographer — hidden, undefined, ephemeral or off-limits — all with “the power to provoke or disorient”. As we tour the half-built edifices of the Archaeological Park of Sicilian Incompletion, the Aralqum Desert (once the Aral Sea), a UK traffic island and the no-man’s-lands between borders, Bonnett reveals how such journeys into dislocation can jar us into awareness of a profound human urge: the love of place and its shadow, the need to escape.

Both are now being eroded by the intensive digital mapping of the planet, which puts at risk the secret and private and makes getting lost a lost cause indeed. This trend, he argues, is destroying our mental landscape: if everything is pinned down and chewed over, where are the mysteries that ignite our urge for discovery? And is cyberspace, as some now claim, more beguiling than terra firma? Like Bonnett (and Huck Finn), I’d rather “light out for the territory”.

Barbara Kiser is Nature’s Books and Arts editor.

Physics in Mind: A Quantum View of the Brain

WERNER R. LOEWENSTEIN
Basic: 2013.

In the delightful and challenging yet accessible *Physics in Mind*, Werner R. Loewenstein unpicks the brain through a physicist’s eyes. He reveals how evolution pushed molecules to become information scavengers and eventually built the neuron network that culminates in consciousness. Working up from basic principles, he takes physics concepts such as quantum theory and applies them to physiology, the senses and neuroscience. Along the way, he offers engaging insights into the parallels between these very different worlds.

After some intriguing diversions, such as one on the mechanics of gut feelings, Loewenstein concludes with the hypothesis that on at least one level, the brain could be a kind of quantum computer. It is a leap. But it is a relatively persuasive one, based on known examples of biological quantum phenomena (such as the coherent quantum waves that form when photoreceptor proteins in the eye respond to light) and similarities between parallel processing in the mind and in quantum computing. Readers hankering for a quantum answer to the mystery of consciousness will not find one here. But they will see how the brain performs the remarkable feat of extracting meaning from its environment.

Elizabeth Gibney is a reporter for Nature in London.

Infinitesimal: How a Dangerous Mathematical Theory Shaped the Modern World

AMIR ALEXANDER
Oneworld: 2014.

Since antiquity, great minds had struggled with a paradox: a line is the sum of its points, but each point has zero length, so how can the line have length? Galileo and other seventeenth-century scientists favoured a radical solution: that geometric objects are made of units whose size is ‘infinitesimally’ small but still not zero.

This seemed to fly in the face of the accepted rationalist tradition of Euclid, in which incontrovertible truths are deduced from a few basic, ‘self-evident’ axioms. And, historian Amir Alexander argues in this original book, the asymptotic fudge challenged the authority of the Catholic Church. The Jesuits had embedded Euclidean geometry into a hierarchical system of knowledge to help block the tide of Protestantism and re-establish Rome’s ideological hegemony.

As Alexander shows, Jesuit hostility and the order’s authority over the Inquisition destroyed the Renaissance spirit of inquiry. Recall, the Inquisition found Galileo guilty of heresy over his advocacy of a heliocentric Universe. Italy became a scientific backwater just as the ‘new science’, founded by Newton on infinitesimal calculus, began to flourish in England. Alexander does not discuss the nineteenth-century creation of rigorous foundations for calculus, which might leave some thinking the field was built on riddles. But his book is a triumph.

Davide Castelvecchi is Nature’s online news editor.

The Compatibility Gene: How Our Bodies Fight Disease, Attract Others, and Define Our Selves

DANIEL M. DAVIS
Allen Lane/Oxford University Press: 2013.

There is a case to be made that the last frontier of knowledge is not the edge of the Universe, or the world of the inaccessibly small, but something much closer to home: our immune system. For all its influence on nearly everything we do, much of it remains a mystery. And where popular-science shelves groan with books about cosmology and quantum mechanics, or evolution and dinosaurs, accessible books on immunology are almost non-existent.

Daniel Davis’s *The Compatibility Gene* goes a long way towards filling that void. Davis relates, with brio, the story of the early days of organ-transplant experiments. He shows how at root, it is thanks to the immune system that the body tells the difference between itself and foreign invaders — with implications for everything from the success of transplants and blood donation to the structure of society and choice of mate. It is not perfect, occasionally straying into hyperbole, but Davis gets a gold star simply for trying, and for putting over an arcane subject with such infectious enthusiasm. (Did I say ‘infectious’?).

Henry Gee is a senior biology editor for Nature.

From Pompeii: The Afterlife of a Roman Town

INGRID D. ROWLAND
Belknap: 2014.

From Pompeii is very much a tale of two cities — one preserved by Vesuvius's eruption in AD 79, the other the modern urban development beyond the gates of the remains. Eschewing the classic historical approach, Ingrid Rowland instead picks up the story once the dust has settled. Sifting through the 1,900 years since the ancient city's demise, she reveals the story of its rediscovery, the influence of its ruins on the cultural development of the modern landscape and, crucially, the problems faced by the artefacts and buildings that have slowly been brought to light.

Marshalling a stellar cast, including Mozart, Dickens, Renoir, Mark Twain and Ingrid Bergman, Rowland is ever engaging and offers diverting anecdotes. Nineteenth-century tourists, for instance, 'gassed' dogs by holding them over emanations of carbon dioxide rising from fissures in the Grotta del Cane.

The modern life of the ruins and town get equal billing. But the ancient site gets the last word, tempered by the growing realization of just how fragile Pompeii's remains are — a significant proportion of its paintings have already been lost to the elements. The site, with so many of its treasures under seemingly permanent lock and key, could now be reaching the end of its inspirational afterlife.

Colin Sullivan is Nature's *chief subeditor*.

Correspondence

Breast-screening trials are ethical

The ethics have been questioned of running new randomized controlled trials to determine the benefits and possible harms of population screening for breast cancer (see go.nature.com/vw13jv). As an ethics representative on the Swiss Medical Board, I believe that there is a moral requirement for this type of study. We need to ascertain whether advances in treatment have cancelled out the benefits of early diagnosis through screening.

The random allocation of women into groups that have mammograms with different detection thresholds (see H. G. Welch *The New York Times* 29 December 2013; go.nature.com/scyt6b) or no mammogram would be ethically problematic if we knew that screening provided a significant net benefit. But this has not been established.

I contend that most women would prefer to participate in a trial that helps to clarify the benefit of screening, rather than continuing to be subjected to screening of doubtful benefit and with potential for significant harm through overdiagnosis.

Nikola Biller-Andorno
University of Zurich, Switzerland.
biller-andorno@ethik.uzh.ch

Pollination curbs climate risk to cocoa

Cocoa yields will be 75,000 tonnes short of those predicted for the 2013–14 growing season (see go.nature.com/ob6eyt). Supply deficits have been linked to extreme weather events in the cocoa-growing regions of Côte d'Ivoire, Ghana and Indonesia, so future production strategies should take climate change into account (see P. Läderach *et al. Clim. Change* **119**, 841–854; 2013).

The global trade in cocoa is worth US\$10 billion per year,

and 50 million people depend on cocoa farming. Farmers are under pressure to stabilize yields, often by increasing chemical inputs and removing shade trees to increase the productivity of cocoa trees. But both strategies could kill vital pollinators.

Farmers need to adopt more sustainable ecological methods. For example, shade trees can make plantations more resilient to drought and provide habitat for pollinators. Increasing insect pollination by as little as 10% can double cocoa yields (see J. H. Groeneveld *et al. Perspect. Plant Ecol. Evol. Syst.* **12**, 183–191; 2010).

Government and industry sustainability initiatives akin to those run by the World Cocoa Foundation (see go.nature.com/uj9nfh) could contribute to cocoa-pollination research and advise farmers on how best to meet the world's increasing demand for chocolate.

Thomas C. Wanger* *University of Göttingen, Germany.*
tomcwanger@gmail.com

**On behalf of 4 correspondents (see go.nature.com/wh86eh for full list).*

Citation databases omit local journals

South America's research impact is underestimated in the main citation databases, as you suggest (*Nature* **510**, 202–203; 2014). However, incorporating the Scientific Electronic Library Online (SciELO) Index into Thomson Reuter's Science Citation Index will not rectify the situation.

SciELO captures only a fraction of the continent's peer-reviewed publications. For example, it indexes papers from just 267 of Brazil's 1,909 journals. Coverage by Elsevier's Scopus database is also inadequate.

According to Latin America's most comprehensive database on scholarly journals, the Latindex Catalog, 4,882 journals in South America meet specified editorial

criteria (see go.nature.com/lbrng2). Scopus includes only 726 of these journals (15%); regional journals are not eligible (see also go.nature.com/laywal).

That leaves 4,156 journals whose impact is hidden from Scopus. If we assume conservatively that each of these publishes only 20 articles per year by South American authors, then at least 83,120 articles are being overlooked annually.

The commercial databases should close this coverage gap to properly reflect the impact of research in South America.

Juan Pablo Alperin *Stanford University, California, USA.*
juan@alperin.ca

Forest devastated by mining is reborn

Open-pit gold mining leaves millions of hectares of wasteland, particularly in the tropics. A series of affordable, socially inclusive and ecologically sound forest-restoration projects in Colombia could become a model for rescuing mined lands around the world.

Reforestation began in 2002 across 1,290 hectares in Cáceres, one of Colombia's most conflict-ridden regions (see L. G. Moscoso Higueta *Reforestation: A Natural Process*; Editorial Colina, 2005). Projects using similar techniques have since begun in other areas of the country.

First the barren landscape is reshaped using a bulldozer, and the soil is enriched with composted sewage sludge, benign microorganisms and other nutrients. Next, *Acacia mangium* trees are planted (because of their hardiness, fast growth and ability to improve soil by fixing nitrogen and providing abundant leaf litter), along with 10–20 native tree species.

After ten years, the *A. mangium* trees are logged and replaced with well-adapted native species to encourage diversity. The Cáceres site now contains more than 120 different

native tree species and harbours an impressive range of wildlife, including jaguars, sloths and several species of primate.

Local people are involved in all steps of the restoration process, and they share in the social and economic benefits generated, such as increased employment and proceeds from timber sales and carbon credits. These will more than compensate for the initial investment of US\$3,000 per hectare.

Evert Thomas *Bioversity International, Cali, Colombia.*
e.thomas@cgiar.org

Graduate admissions test has some merit

We would like to clarify a couple of points raised by Casey Miller and Keivan Stassun in their criticism of the US Graduate Record Examination (GRE) for selection of graduate students (*Nature* **510**, 303–304; 2014).

As documented in a meta-analysis by independent researchers, the GRE predicts more than just first-year graduate grades: it also predicts higher-degree attainment, time to complete, research productivity and citation counts (see N. R. Kuncel *et al. Psychol. Bull.* **127**, 162–181; 2001).

And although we acknowledge that the GRE “reflects certain demographic characteristics of test-takers”, we reject any implication that this is a fault in the test: rather, test scores reflect the reality that more educational resources are available to students from wealthier families.

We agree that graduate admissions should be based on multiple sources of information, including standardized test scores, undergraduate grades, diligence and non-cognitive factors such as ‘grit’.

Brent Bridgeman, David Payne, Jacqueline Briel *Educational Testing Service, Princeton, New Jersey, USA.*
bbridgeman@ets.org

T_H2 and T_{reg} candidate genes in elephant shark

ARISING FROM B. Venkatesh *et al. Nature* **505**, 174–179 (2014)

Recently in this journal, Venkatesh and co-workers published their analysis of the elephant shark (*Callorhynchus milii*, a chimaera) genome sequence¹. Their analyses suggested that at the evolutionary level of cartilaginous fish, only a primordial T-helper-cell (T_H-cell) system was established, geared towards a T_H1-type response, because they could not find several important candidate genes, or gene features, necessary for other types of T_H cells. However, in contrast to Venkatesh and co-workers, I am able to find elephant shark candidate genes for interleukin-2 (IL-2), IL-4, IL-5 and IL-13, which have important roles in mammalian T_H2 and T-regulatory-cell (T_{reg}-cell) systems; furthermore, I feel that claims by the authors of absent CD4 and FOXP3 functions in cartilaginous fish are insufficiently substantiated. Therefore, I believe that major conclusions by the authors regarding proposed unique features of the T-cell system in the elephant shark are based on false arguments, or, as in the case of their CD4 model, are premature. There is a Reply to this Brief Communication Arising by Venkatesh, B. *et al. Nature* **511**, <http://dx.doi.org/10.1038/nature13447> (2014).

Identities of the cytokine genes reported here are supported by genomic locations (Fig. 1), software predictions of genes, typical intron–exon organizations, in most cases by best matches upon database-similarity searches, and by software predictions of protein structure (Appendix).

In terms of candidate genes for T_H2 cytokines related to mammalian IL-4 and IL-13, between the elephant shark genes *RAD50* and *KIF3A*, there are three putative genes of the IL-4 and IL-13 T_H2 cytokine family, which I subsequently name *IL-4/13A*, *IL-4/13B* and *IL-4/13C* (Fig. 1 and Appendix). The designation ‘4/13’ derives from the fact that the tandem gene duplication resulting in *IL-4* and *IL-13* may have occurred only in the tetrapod line². All IL-4 and IL-13 family members may have T_H2 functions, as suggested by accumulating evidence from studies in fish, and by the discovery of elephant shark IL-4- and IL-13-receptor chain genes by Venkatesh and co-workers.

With regard to candidate genes for T_H2 cytokines related to mammalian IL-5, upstream of elephant shark *RAD50* two putative cytokine genes are located in a similar position to mammalian IL-5 (Fig. 1 and Appendix; the most characteristic motif, encoded by the exon 2–exon 3 border region, is I/LXXPPXXD/E/NXXXCXXI in which a solidus represents ‘or’ and an ‘X’ denotes any amino acid). Orthologous relationships of elephant shark IL-5 (or possibly IL-5-family) candidate genes, which are referred to here as *IL-5A* and *IL-5B*, with mammalian

genes are not clear-cut, and the evolution of the IL-3, IL-5 and CSF2 cytokine family will need a more thorough analysis. However, detection of *IL-5A* and *IL-5B* combined with the detection of a candidate gene for IL-5R α by Venkatesh and co-workers make it likely that IL-5-like functions are present in cartilaginous fish.

In terms of the candidate gene for IL-2, which in mammals is important in the T_{reg} system, in contrast to what Venkatesh and co-workers reported, elephant shark does have IL-2 and IL-21 candidate genes at the expected genomic location (Fig. 1 and Appendix). However, Venkatesh and co-workers seem to be correct that the species lacks *IL-2R α* . Also, in bony fish only a gene with *IL-15R α* features is found³, and *IL-2R α* may have been rapidly and extensively crafted from an *IL-15R α* duplicate starting early in tetrapod evolution⁴. In bony fish both IL-2 and IL-15 bind to IL-15R α (ref. 3), and it should be recognized that T_{reg} function should be possible with only a single IL-15R α -type molecule in the absence of IL-2R α ⁴. Like other vertebrates, besides IL-15 (Appendix; this sequence was already reported by Venkatesh and co-workers), elephant shark has a gene for IL-15-like (IL-15L; Appendix), a protein with unknown function that also binds IL-15R α and that also might participate in functions mediated by IL-2 and/or IL-15 (ref. 4).

For the important negative conclusion by Venkatesh and co-workers that elephant shark does not have a CD4 equivalent, analysis of the syntenic genomic region would be necessary. It is also unclear why CD4 should be more necessary for other T_H subsets than for T_H1, or what a ‘bona fide’ CD4 molecule should look like.

The structural hallmarks of elephant shark FOXP3 that Venkatesh and co-workers suggest to be incompatible with T_{reg} function were shown in a paper that they referenced⁵ to still allow substantial T_{reg} activity in mammals, and extrapolation across vast species borders of the partial reductions in activity described in that paper⁵ should be done with care.

In short, the sequences available so far do not suggest that the elephant shark T_H system is geared towards T_H1 or that it represents a primordial state.

Johannes M. Dijkstra¹

¹Institute for Comprehensive Medical Science, Fujita Health University, Dengaku-gakubo 1-98, 470-1192 Toyoake, Aichi, Japan.
email: Dijkstra@fujita-hu.ac.jp

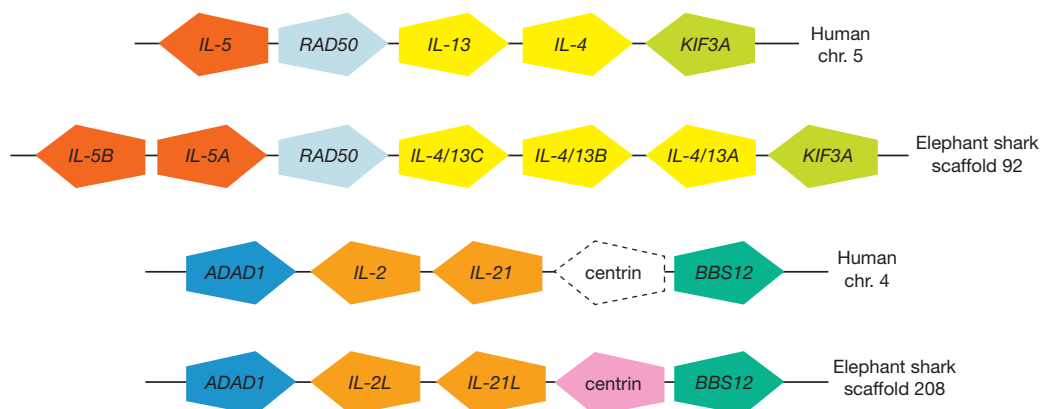


Figure 1 | Genomic location of several cytokine genes. The paper by Venkatesh and co-workers described absence of the elephant shark interleukin

gene candidates that are demonstrated here. The centrin gene in human is a pseudogene.

Received 2 February; accepted 30 April 2014.

1. Venkatesh, B. *et al.* Elephant shark genome provides unique insights into gnathostome evolution. *Nature* **505**, 174–179 (2014).
2. Ohtani, M., Hayashi, N., Hashimoto, K., Nakanishi, T. & Dijkstra, J. M. Comprehensive clarification of two paralogous interleukin 4/13 loci in teleost fish. *Immunogenetics* **60**, 383–397 (2008).
3. Wen, Y., Fang, W., Xiang, L. X., Pan, R. L. & Shao, J. Z. Identification of Treg-like cells in *Tetraodon*: insight into the origin of regulatory T subsets during early vertebrate evolution. *Cell. Mol. Life Sci.* **68**, 2615–2626 (2011).
4. Dijkstra, J. M. *et al.* Identification of a gene for an ancient cytokine, interleukin 15-like, in mammals; interleukins 2 and 15 co-evolved with this third family member, all sharing binding motifs for IL-15R α . *Immunogenetics* **66**, 93–103 (2014).
5. Andersen, K. G., Nissen, J. K. & Betz, A. G. Comparative genomics reveals key gain-of-function events in Foxp3 during regulatory T cell evolution. *Front. Immunol.* **3**, 113 (2012).

Competing Financial Interests Declared none.

doi:10.1038/nature13446

Appendix

Elephant shark cytokine sequences

This Appendix lists the sequences of the predicted elephant shark cytokines discussed in the Comment. Genes of all of the presented sequences have exon–intron organizations that appear similar to, or derived from, the basic organization of the short-helix type I cytokine family, which consists of four phase-0-organized exons encoding: exon 1, leader peptide and α -helix A; exon 2, the loop between α -helices A and B; exon 3, α -helices B and C; exon 4, α -helix D. Genes without messenger RNA reports were predicted from genomic DNA sequences using GENSCAN (<http://genes.mit.edu/GENSCAN.html>) or FGENESH (<http://linux1.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind>) computer software, but in some instances also by hand using comparisons with family members; all presented individual exons were predicted by at least one of the programs GENSCAN and FGENESH. The deduced molecules are indicated per matching exon, with the intron phase given in brackets. Predicted leader peptides identified using SignalP software (<http://www.cbs.dtu.dk/services/SignalP/>) are underlined. For prediction of secondary structures and recognition of homology or analogy with known molecule structures Phyre² software (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>) was used. Best matches upon database similarity searches are shown to help identification; note that this family of cytokines is so highly diverged that often non-related molecules are found as the best match (see also refs 2 and 4).

The elephant shark sequences presented here are available through the Elephant Shark Genome Project (<http://esharkgenome.imcb.a-star.edu.sg/>), and NCBI (GenBank) databases that were used can be found at <http://www.ncbi.nlm.nih.gov/>.

Elephant shark IL-4/13A

The open reading frame is situated in *C. milii* scaffold 92 (KI635946) between positions 1433512 and 1435606, in reverse orientation. The deduced amino acid sequence is:

MSKAVSIMIVLMTVFARSDCGVLREILSKRQAMGELIQTLLQLTAGE (0)

VACGNTMVDADVVFAGN (0)

AAPESFCKAARALLQAPACKPIEQQLNNLIINLQITGVE (0)

TSNCSLDENSEVEMSTFLISLKKFKCKVFSLQ

- Best match in NCBI database upon ‘blastp’ analysis: ADH51720, IL-4 of camel.
- Best match using the Phyre² analysis: d2b8ua1, IL-4, confidence, 97%.

Elephant shark IL-4/13B

The open reading frame is situated in *C. milii* scaffold 92 (KI635946) between positions 1427088 and 1429667, in forward orientation. The deduced amino acid sequence is:

MAHILKAAALLSILLSAVCATQITTKARHQSLKEISTIRHILENQ (0)

VHHHQNHVPEIFEDVK (0)

DRSDSEVFCKAAQVLELLPFYNNLKLHKLRLHNLHMGQ (0)

ITECEVSELSQVEFKNFLKRLNFSQKKFRTNGDKNHLTTKSKG

- Best match in NCBI database upon ‘blastp’ analysis: XP_006023629, IL-4 of alligator.
- Best match by Phyre² analysis: d2b8ua1, IL-4, confidence, 100%.

Elephant shark IL-4/13C

The open reading frame is situated in *C. milii* scaffold 92 (KI635946) between positions 1414852 and 1416816, in forward orientation. The deduced amino acid sequence is:

MKPLTIAFMFLVLTFSEGCKCRAVTEITDLGFLEK (0)

HMAKLVPVIWNENIR (0)

PSNGSVSCKSHDVLRLKTVLNRSQFVAKRTEQLIHNLAVGSV (0)

NQTCVLNARSEIDLPAALLSKLNEMRRRL

- Best match in NCBI database upon ‘blastp’ analysis: a non-related protein.
- Interleukin best match in NCBI database upon ‘blastp’ analysis: AAL18820, IL-4 of cotton rat.
- Best matches by Phyre² analysis: c2kka, ataxin-7 (a non-related molecule), confidence, 66%; and d2b8ua1, IL-4, confidence, 31%.

Elephant shark IL-5 (or IL-5-family) candidate IL-5A

The open reading frame is situated in *C. milii* scaffold 92 (KI635946) between positions 1330410 and 1333842, in forward orientation. The deduced amino acid sequence is:

MTKMICVYVLLFNSFLFVGRNTHPCSRPQIYETINAIKKHEQDIIK (0)

KLPAELESPINVD (0)

DENCKQQINSTFYKGLKPLYETDLKPIAEELSNMTNDKADVQ (0)

DKQSDCNQNVQTVNGAKFLKRLNLYFSKMSRMSRRRRSLKQI LANFH

- Best match in NCBI database upon ‘blastp’ analysis: a non-related protein.
- Interleukin best match in NCBI database upon ‘blastp’ analysis: XP_004686760, IL-5 of mole.
- Best matches by Phyre² analysis: c3c2gB, sys-1 protein (a non-related molecule), confidence, 6%; and d2psma1, IL-15, confidence, 20%.

Elephant shark IL-5 (or IL-5-family) candidate IL-5B

The open reading frame is situated in *C. milii* scaffold 92 (KI635946) between positions 1322789 and 1325410, in reverse orientation. The deduced amino acid sequence is:

MISFYMLLFASSLFTLGSRKPCHYSTICEMLKELKKVEKL (0)

VLPGNLQAPINVE (0)

NKQCVQTHIGFNRGFKQLKKELHKRDPTTSDIFCIILSNLHYEEM HPK (0)

LPFQNKHSSCTASDINVEEMNSAKFLKKFKNFLQAYYNTLLNQNKERK

- Best match in NCBI database upon ‘blastp’ analysis: a non-related protein.
- Interleukin best match in NCBI database upon ‘blastp’ analysis: XP_005510723, IL-21 of pigeon.
- Best match by Phyre² analysis: d2psma1, IL-15, confidence, 93%.

Elephant shark IL-2-like (IL-2L)

The open reading frame is situated in *C. milii* scaffold 208 (KI636062) between positions 426689 and 428986, in reverse orientation. The deduced amino acid sequence is:

MPKLLCLFFTVLFTTAIGDDGTFKFYLLREMILKTKELQTLLK (0)
NDQYLMNSQDLIRTPVILE (0)

PNCSLICFEEGLTTLKHCKKLPSCCKPASIIISNMKTFQTV (0)

EGSQRCFEFKQKKIPEFLALKVFLQRNYKTAQKHVSSTGMKSKPYV

- Best match in NCBI database upon 'blastp' analysis: ABS44960, IL-2 of pufferfish.
- Best matches by Phyre² analysis (in order, starting with best match): d2z3qa1, IL-15, confidence, 99%; d2psma1, IL-15, confidence, 99%; d2oqpa1, IL-21, confidence, 99%; d1m47a, IL-2, confidence, 80%.

Elephant shark IL-21-like (IL-21L)

The open reading frame is situated in *C. milii* scaffold 208 (KI636062) between positions 460264 and 470256, in reverse orientation. The deduced amino acid sequence is:

MNRIIFYSFILAAFI AVLHAAPRRDHVVLKEMMKDVRNLKKMFG (0)

QSGVTVRTTESKL (0)

DNLSCYMAELKGLKKSRLMQINTVEKLNKRAKQLNLLN (0)

IDCPNSEN YMKN SLEQFLNQFYVHLQKKFHA AVK (0)

VGQQLNKVQVNI

- Best match in NCBI database upon 'blastp' analysis: a non-related protein.
- Interleukin best match in NCBI database upon 'blastp' analysis: XP_006010899, IL-21 of coelacanth (this seems to be erroneously annotated as 'IL-15-like' in the GenBank accession).
- Best matches by Phyre² analysis (in order, starting with best match): d2psma1, IL-15, confidence, 97%; d2z3qa1, IL-15, confidence, 96%; d2oqpa1, IL-21, confidence, 94%.

Elephant shark IL-15

The open reading frame is situated in *C. milii* scaffold 57 (KI635911) between positions 623715 and 646704, in reverse orientation. The

sequence is also supported by the mRNA sequence of GenBank JW878023. The deduced amino acid sequence is:

MQFAVFKQ (0)

RHVKRVCFCCLYNLTLDVFTFINTEAGLCCLLLC (2)

SLSVCLPKVEGDSTTKLDKKHGMREMMSTLMNLN (0)

KSLSIPASIS (0)

DLTLYSPENVP (0)

ENCFDAALLCFMKEMKTLEFEMKLVQSKNHQKLSICNQMLTTY
KRLHE (0)

LNDCKPCEEFEQKNVTSFLSGFISLLQIIKFKSVN

- Best match in NCBI database upon 'blastp' analysis: AGQ17908, IL-15 of nurse shark.
- Best matches by Phyre² analysis (in order, starting with best match): d2psma1, IL-15, confidence, 100%; d2z3qa1, IL-15, confidence, 100%; d2oqpa1, IL-21, confidence, 99.9%.

Elephant shark IL-15-like (IL-15L)

Some of the exons are situated in the short *C. milii* scaffold 5514 (KI639395) in reverse orientation. The full-length sequence is encoded by the mRNA sequence of GenBank KA353649. The deduced amino acid sequence is:

MRLWFSVS SVYLV MGLVSAKVKKNDCAKECFGVFFKLQEIV† (0)

RTEEDIRLYTPHGQTAK (0)

VCPQKTTVCFAAELHVLIFEYDGIWKENVTELAASVIECLEHFL
HNL† (0)

RMPSACGHTEPSLPTGCPMCEQFHEESVLLFLEHLKKFLQFIYYC
CNICLSE

†Sequences for which genomic information is not available.

- Best match in NCBI database upon 'blastp' analysis: NP_001009558, IL-15L of zebrafish.
- Best matches by Phyre² analysis (in order, starting with best match): d2z3qa1, IL-15, confidence, 99.6%; d2psma1, IL-15, confidence, 99.6%; d2oqpa1, IL-21, confidence, 99.2%

Venkatesh *et al.* reply

REPLYING TO J. M. Dijkstra *Nature* **511**, <http://dx.doi.org/10.1038/nature13446> (2014)

In the accompanying Comment¹, Dijkstra has identified gene models potentially encoding cytokines in regions of the *Callorhinchus milii* genome, whose syntenic counterparts in teleost and tetrapod genomes harbour interleukin genes. These genes were not predicted by our assembly and annotation pipeline². TBLASTN searches of the genome assembly and RNA-sequencing (RNA-seq) transcriptomes from 10 tissues (including thymus and spleen) of the elephant shark, and RNA-seq transcriptomes from the thymus and spleen of nurse shark using human and teleost fish interleukins as query sequences, were also unable to pick up these genes; although we readily detected other cytokine genes in the same family (for example, interleukin-7 (IL-7) and IL-15). Below are our comments on the candidate genes identified by Dijkstra.

With regard to candidate genes for IL-4 and IL-13, of the three genes (IL-4/13A, IL-4/13B and IL-4/13C), gene IL-4/13C lacks readily identifiable canonical features of IL-4 and IL-13, whereas genes IL-4/13A and IL-4/13B exhibit weak sequence similarity to their homologues in other vertebrates (Fig. 1). It is possible that these genes encode primordial versions of canonical IL-4 and IL-13 cytokines. Interestingly, a single exon shuffling event—a process implicated in the birth of new genes³,

replacing exon 3 of gene IL-4/13A with the corresponding exon of gene IL-4/13B—could give rise to a hybrid gene encoding sequence signatures and disulphide bonds typical of tetrapod IL-4 and IL-13 proteins⁴ (Fig. 1). This scenario is consistent with the presence of a single IL-4/13-like gene in the corresponding genomic region of a basally branching ray-finned fish, the spotted gar (http://www.ensembl.org/Lepisosteus_oculatus/Info/Index).

In terms of candidate genes for IL-5, the two putative *C. milii* IL-5 proteins identified by Dijkstra have no readily identifiable similarity to other known IL-5 proteins.

In contrast to the constitutive, high expression levels of canonical type 2 T-helper (T_H2) cytokines in the thymus and gills of teleosts⁵, analysis of the RNA-seq reads of *C. milii* indicated that the above five candidate 'T_H2' genes are either not expressed (IL-4/13C and IL-5A) or expressed at very low levels (1.12, 1.42 and 0.88 transcripts per million for IL-4/13A, IL-4/13B and IL-5B, respectively) in the gills, when compared to the homeostatic interleukins IL-15 and IL-15L (14.37 and 48.27 transcripts per million).

©2014 Macmillan Publishers Limited. All rights reserved

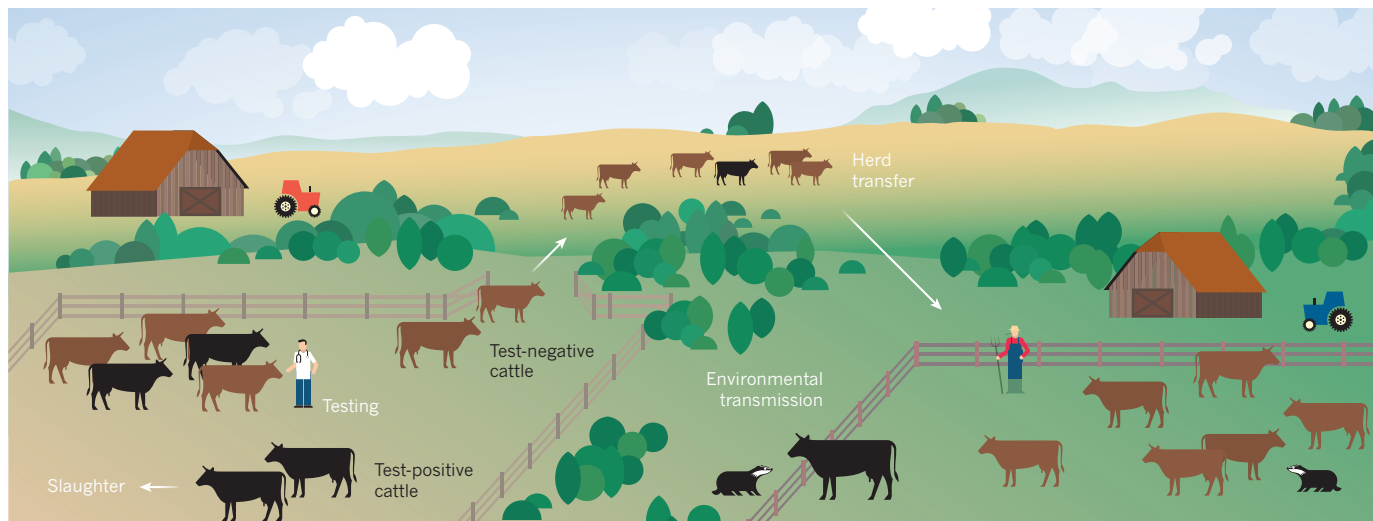


Figure 1 | Bovine tuberculosis spread. Brooks-Pollock *et al.*⁶ present a model that parameterizes the complex processes involved in the spread of bovine tuberculosis in Great Britain, including cattle movement, cattle testing (incorporating problems with current diagnostics that miss true positives), environmental exposure and control strategies for cattle and wildlife. Among other findings, the model suggests that a minority of farms act as superspreaders of infection and that control of badger populations is unlikely to stem the growth of the current epidemic.

ANIMAL HEALTH

How to control bovine tuberculosis

A model of the transmission and spread of bovine tuberculosis in Britain suggests that controlling the epidemic will require large-scale cattle slaughter or a major rethink of combined control strategies. [SEE LETTER P.228](#)

ROBBIE A. MCDONALD

Bovine tuberculosis is a huge economic and social problem in Britain, with control measures costing taxpayers about £100 million (US\$170 million) per year. Control is technically problematic and politically controversial because of the involvement of wild badgers in the maintenance and transmission of the infection¹. Badgers are alternately much-loved and loathed in British society² and badger culling has complex positive and negative outcomes for disease in cattle^{3–5}. On page 228 of this issue, Brooks-Pollock *et al.*⁶ describe a dynamic model of bovine tuberculosis transmission within and between farms that is unprecedented in its scale, realism and approach. By taking bovine tuberculosis by the horns, they have provided an instructive but ultimately disheartening insight into prospects for controlling this chronic disease.

Brooks-Pollock and colleagues modelled the entire national cattle herd, using records of the daily movements of 30,000 cattle between farms and real protocols for bovine tuberculosis (bTB) testing and movement controls. Their approach was computationally intensive, it acknowledged uncertainty

and it used real-life evidence of bTB incidents in cattle herds and numbers of infected cattle to ‘reverse engineer’ the parameters of transmission that could give rise to the observed epidemic (Fig. 1).

The authors also scrutinized cattle-testing protocols that rely on a skin test with a known tendency to miss true positives⁷. They highlight the fact that reliable detection in a herd depends on there being multiple infected individuals and that there is a high probability of infected animals being left behind in the herd, from which subsequent cases can be seeded. Thus, clearing infection from herds is extremely problematic. However, Brooks-Pollock and colleagues worked within this diagnostic constraint by distinguishing between infected herds and herds in which infection had been detected, allowing them to model the consequences of missing and moving infected cattle. They also assessed the importance of infection from the local environment, comprising direct transmission from wildlife (primarily badgers) and indirect exposure to environmental contamination stemming from cattle and/or wildlife.

The model reveals that movement of infected animals was responsible for 84% of

newly infected farms and, crucially, that there was a heavy bias towards as few as 10% of farms, characterized by selling many animals, as the source of nearly all secondary cases. Similarly, a small minority of farms seemed to be responsible for spread through the environment. Thus, a key finding of the study is that a small proportion of farms probably function as superspreaders of bTB infection.

The causes of incidents of infection in a herd were evenly spread across three principal processes: introduction of an infected animal; transmission from the local environment, including from wildlife; and an infected animal being left behind after testing and slaughter of test-positive animals. No single cause dominated, so it comes as little surprise that when the authors tested potential control treatments in their model, they found single, discrete strategies to be generally ineffective in checking the growth of the epidemic.

In fact, only one of ten treatments — slaughtering the whole herd upon detection of bTB — was effective in rapidly bringing all measures of the epidemic well below the baseline. The model suggests that this treatment would result, in the year after implementation, in drastic reductions in the number of cattle

testing positive for bTB and in removal of restrictions from much of the national herd. However, this is a silver bullet that would cause severe wounding to the cattle industry, because it would come at the cost of a one-off, 20-fold increase in cattle slaughtered in the first year of the policy. Such a cost is reminiscent of the catastrophic scale of Britain's 2001 experience of controlling foot-and-mouth disease and is likely to be horrifying to policy-makers, farmers and animal-welfare campaigners alike. Faced with the current epidemic trajectory, however, the authors suggest that it might be judicious to contemplate how such severity trades off against ongoing, large-scale culling well into future decades.

The model also revealed that a one-off test of the national cattle herd would bring success in reducing all measures of disease some three to four years later. This treatment was implemented in Wales in 2008–09, and the model's predictions are consistent with current tentative observations of bTB decline in Wales. The model further predicts that improved cattle-test performance would help to reduce trade restrictions, but at a cost of higher numbers of cattle culled, and that cattle vaccination offers some encouraging prospects, although not at the usual level expected for mass vaccination campaigns. Disappointingly for supporters of England's current pilot badger culls, an across-the-board 50% reduction of all sources of environmental transmission, simulating successful culling or vaccination of badgers, was shown by the model to have little impact on any measures of cattle disease, and failed to prevent ongoing growth of the epidemic.

Brooks-Pollock and colleagues' complex modelling approach therefore provides two simple messages: first, that direct, generic controls imposed on cattle provide the greatest leverage for acting on the current mass of the bTB epidemic; and second, that although the environment is an important source of infection, indirectly tackling cattle disease by managing wildlife, by whatever means, is likely to yield unimpressive results and contribute little biologically to controlling a national epidemic.

Although their work is compelling, the authors' job is not complete. Policy-makers often speak of their use of a 'package of measures' and 'every tool in the box' in controlling bTB, particularly with respect to controversial policies to cull badgers. Measures aimed at wildlife could contribute to bTB control in Britain, perhaps as part of achieving consensus among stakeholders in conflict or, most importantly, in the eventual transition from control to eradication of infection. But having shown that individual strategies, short of draconian whole-herd culling, will not rapidly turn this epidemic around, Brooks-Pollock *et al.* now have the perfect opportunity to use their model to test policy-makers' expectations of their various packages and toolboxes and to gauge their

overall cost-effectiveness. UK governments and the farming industry badly need such guidance to frame and implement their overall strategy and to manage expectations among their diverse stakeholders. ■

Robbie A. McDonald is at the *Environment and Sustainability Institute, University of Exeter, Penryn TR10 9FE, UK.*
e-mail: r.mcdonald@exeter.ac.uk
Twitter: [@robbieamcdonald](https://twitter.com/robbieamcdonald)

APPLIED PHYSICS

A new view on displays

Materials that rapidly switch between amorphous and crystalline states are widely used to manage heat and store data. They now emerge as promising building blocks for ultrahigh-resolution display devices. [SEE LETTER P.206](#)

DIRK J. BROER

A modern world without electronic display devices is unthinkable. In rich countries, most households have several such devices: television sets, mobile phones and tablets, to name a few. Alongside trends to make large displays (home cinemas and billboards), there is a growing interest in creating ultrasmall displays with high resolution and fast response times — the amount of time a pixel in the display takes to change. Such microdisplays are already used in technologies such as microprojectors and wearable devices. On page 206 of this issue, Hosseini *et al.*¹ report how they take up this trend to go ultrasmall by developing an optoelectronic device that could be used to make ultrathin displays of ultrahigh resolution.

At the heart of Hosseini and colleagues' device is a film of a germanium antimony tellurium alloy ($\text{Ge}_2\text{Sb}_2\text{Te}_5$; GST), which is classed as a phase-change material. Phase-change materials switch rapidly between a metastable amorphous phase and a stable crystalline phase (Fig. 1a, b) — both of which exist at room temperature^{2,3} — in response to an external stimulus, such as heat, light or electrical current. This last stimulus, which induces GST to undergo the phase change, is most relevant for display applications. Phase-change materials are mostly known for their use in thermal-energy management; they store or release heat during the phase transition, and so offer good heat-management solutions. But they can also be used as media for rewritable optical data storage, owing to the difference in optical properties between the two phases⁴. Moreover, they have great potential as non-volatile electronic memories, because the two phases have vastly different electrical conductivities⁵.

To create their optoelectronic device, Hosseini *et al.* sandwiched a layer of GST, a

- Godfray, H. C. J. *et al. Proc. R. Soc. B* **280**, 20131634 (2013).
- Cassidy, A. *Sociol. Rev.* **52**, 192–214 (2012).
- Donnelly, C. A. *et al. Nature* **439**, 843–846 (2006).
- Woodroffe, R. *et al. Proc. Natl Acad. Sci. USA* **103**, 14713–14717 (2006).
- Bielby, J., Donnelly, C. A., Pope, L. C., Burke, T. & Woodroffe, R. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1401503111> (2014).
- Brooks-Pollock, E., Roberts, G. O. & Keeling, M. J. *Nature* **511**, 228–231 (2014).
- Conlan, A. J. K. *et al. PLoS Comput. Biol.* **8**, e1002730 (2012).

This article was published online on 2 July 2014.

few nanometres thick, between two layers of indium tin oxide, which act as transparent electrodes. They found that the optical transmission and reflection properties of this structure are determined by the thickness of the individual layers and their refractive indices. On application of an appropriately shaped pulse of electrical current between the two electrodes, the GST film can switch between the crystalline and the amorphous states, changing the film's refractive index and its colour. The colour of the film in each state can be chosen by tuning the thickness of one of the transparent electrodes (Fig. 1c), and is visible by illuminating the structure with 'white' (broad-band) light.

Colour changes are a crucial element of display technologies, but to make a display, an array of differently coloured pixels must be fabricated in which each pixel's colour is individually switched. It is this point in particular that makes phase-change materials an attractive option with which to build the pixels of microdisplays. By using lithographic techniques, the pixel sizes can be brought down into the submicrometre region, enabling an ultrahigh resolution that exceeds that of most other existing technologies. In current colour displays, each pixel is individually controlled by a type of electrical switch known as a thin-film transistor. A high-resolution technology based on a GST device such as that demonstrated by Hosseini *et al.* could allow the full incorporation of thin-film transistors, and could be built in a single production line.

Hosseini and colleagues' GST-display principle is different from that of conventional display technologies in many aspects. Their optoelectronic device, from which a future display could be made, is an entirely solid-state system that can be easily manipulated, even on flexible and bendable substrates. Such

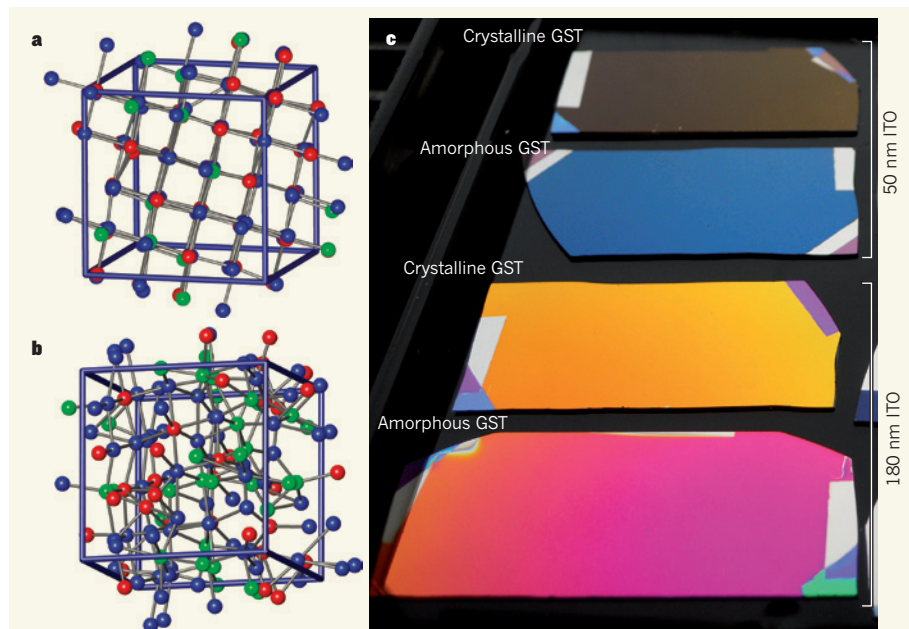


Figure 1 | Inducing colour changes in a germanium antimony tellurium alloy. a, b, Examples of lattice models³ of the crystalline (a) and the amorphous (b) phases of $\text{Ge}_2\text{Sb}_2\text{Te}_5$ (Ge, green; Sb, red; Te, blue). c, Hosseini *et al.*¹ have sandwiched a layer of $\text{Ge}_2\text{Sb}_2\text{Te}_5$ (GST) between two transparent layers of indium tin oxide (ITO) and shown that the material can be electrically switched between its amorphous and crystalline states and undergo colour changes. In the examples shown here, a 7-nanometre-thick GST layer was embedded between a 10-nm-thick top ITO layer and a bottom ITO layer either 180 nm or 50 nm thick. The reflected colour of the layered structure depends on the GST phase and the thickness of the bottom ITO layer. Each structure is approximately 5 centimetres long and 2 cm wide.

a GST display would also provide bistability, a property that allows an image to be maintained over a long period even when the power of the display is switched off or is in standby mode. Bistability is useful to reduce power consumption, for example in electronic-book readers. The authors demonstrated a simple, static display that already shows this feature. Power usage could be further improved by operating the GST display in a mode that provides daylight visibility, even under bright outdoor conditions.

Many applications, such as three-dimensional displays, require a high image-refresh rate — the number of times per second that an image is refreshed — to deliver full-motion video content. The GST device described here is intrinsically ultrafast. In solid-state memory applications, GST films are known to switch between the amorphous and the crystalline phases within 100 nanoseconds⁶. In display applications, this timescale would correspond to an image-refresh rate of 10 megahertz — more than 20,000 times faster than present television technologies.

However, there are several issues that need to be addressed before the proposed technology can enter the display market. First, and as already mentioned, the GST device will need to be pixelated and controlled by thin-film transistors. Second, the device should be able to be connected to electronic components that ensure that each pixel receives the right electrical pulse at the right moment. Third,

the range of possible GST-film colours has to be expanded to fill the whole gamut, which is necessary to create vivid colours. Fourth, the display should also be able to give a fully black

STEM CELLS

The quest for the perfect reprogrammed cell

There are two methods for reprogramming mature cells to pluripotent stem cells, which can give rise to all cells of the body. The first direct comparison of the methods reveals that both can cause subtle molecular defects. [SEE ARTICLE P.177](#)

VLADISLAV KRUPALNIK & JACOB H. HANNA

Pluripotent stem cells hold promise for disease modelling and therapeutics, because they have the potential to differentiate into almost all cell lineages. In particular, there is much interest in patient-derived pluripotent stem cells, which are genetically matched to the patient's own cells, minimizing the risk of rejection by the immune system. In the past decade, two cell-reprogramming methods have been successfully used to generate patient-derived pluripotent stem cells: cloning¹ and direct reprogramming of

or a fully white colour, to allow high-contrast colour images to be obtained. Finally, high-end displays are also judged on their ability to provide a large number of grey scales, which is not an obvious property of phase-change films.

If and when these issues are appropriately addressed, this new display concept may find use in applications that are outside the reach of current display technologies. The optoelectronic structure demonstrated by the authors could, in principle, provide functions other than image display, including sensing of incident light, data storage and performing logic operations. The device's image-display capability could also be used in emerging technologies such as smart contact lenses and other wearable devices, or even in implantable displays in human bodies. ■

Dirk J. Broer is at the Institute for Complex Molecular Systems and in the Laboratory for Functional Organic Materials and Devices, Eindhoven University of Technology, Den Dolech 2, 5612 AZ Eindhoven, the Netherlands.
e-mail: d.broer@tue.nl

1. Hosseini, P., Wright, C. D. & Bhaskaran, H. *Nature* **511**, 206–211 (2014).
2. Yamada, N., Ohno, E., Nishiuchi, K., Akahira, N. & Takao, M. *J. Appl. Phys.* **69**, 2849–2856 (1991).
3. Hegedüs, J. & Elliott, S. R. *Nature Mater.* **7**, 399–405 (2008).
4. Alfonso, C. N., Solis, J., Catalina, F. & Kalpouzos, C. *Appl. Phys. Lett.* **60**, 3123–3125 (1992).
5. Fritzsche, H. *Annu. Rev. Mater. Sci.* **2**, 697–744 (1972).
6. Simpson, R. E. *et al. Nature Nanotechnol.* **6**, 501–505 (2011).

differentiated cells to induced pluripotent stem cells, through the addition of a defined transcription-factor cocktail². However, the molecular differences between cells derived using each method remain unclear. In this issue, Ma *et al.*³ (page 177) define the molecular differences between these two cell types, and set the stage for further improvements to reprogramming methodology.

Derivation of induced pluripotent stem (iPS) cells is an appealing technology, because iPS cells can be reproducibly derived from patient samples⁴. But comparison of iPS cells with the pluripotent embryonic stem (ES)

cells generated during normal embryogenesis shows that human iPS cells are not completely reprogrammed, and reveals epigenetic differences between the two cell types⁵ (epigenetic marks are lingering genomic modifications that affect gene expression without changing DNA sequence). Understanding the biological origins of such differences is necessary if we are to continue to optimize iPS-cell reprogramming.

Cloning—also called somatic cell nuclear transfer (SCNT)—involves transfer of the nuclear material from a mature donor cell into an egg from which the nucleus has been removed. Pluripotent cells, called nuclear transfer ES (NT ES) cells, which are genetically matched to the donor, then arise as the egg begins to develop into an embryo. Generation of patient-specific NT ES cells from adult human cells is now feasible⁶. Although SCNT does not involve introducing transcription factors that have the potential to cause cancer (which is a problem with iPS cell generation), the protocol is technically difficult. It also requires donation of eggs, which creates health, safety and ethical challenges.

Ma and colleagues conducted a comprehensive molecular analysis of iPS cells and NT ES cells that were derived from the same mature cell line, and so were genetically matched (Fig. 1). They compared the reprogrammed cells to *in vitro* fertilized (IVF) ES cells⁷—stem-cell lines derived from embryos produced by IVF, which have therefore not undergone reprogramming. The eggs used to generate IVF ES cells and those used for SCNT

were provided by the same person.

The authors first focused on identifying differing large genetic duplications or deletions (known as copy number variations) in the three cell types. On average, they detected two copy number variations per iPS cell line, but this was not significantly higher than in NT ES or IVF ES cells. Furthermore, Ma and co-workers found that up to 30% of iPS cells had no identifiable genetic abnormalities, indicating that copy number variations do not always occur during iPS cell reprogramming, and that such lines can be obtained with reasonable efficiency. Although the occurrence of genetic mutations in individual nucleotides was not analysed in the current study, these results reiterate previous indications⁸ that deep sequencing should be used to ensure that iPS and NT ES cell lines are mutation free.

Next, Ma *et al.* profiled DNA methylation, an epigenetic modification that regulates gene expression patterns⁹. Perfectly reprogrammed ES cells would be expected to carry exactly the same methylation marks as those generated by IVF ES cells. In genome-wide analyses, the authors found that methylation in NT ES cells was more similar to that in iPS cells when compared with IVF ES cells. The iPS cells did not carry any specific type of abnormality that was not found in NT ES cells, but rather retained many more altered methylation patterns.

Finally, Ma and colleagues studied transcriptional differences in the three cell types. They observed that incomplete demethylation patterns correlated with abnormal gene

transcription in iPS cells. Again, NT ES cells were more similar to IVF ES cells, although some transcriptional alterations were apparent in both reprogrammed cell types.

What is the basis for the epigenetic differences between reprogrammed cell types? Compared with iPS cell generation, the reprogramming cues dictated by the egg in SCNT are more natural. This probably explains the relatively more normal DNA-methylation patterns in NT ES cells. Notably, however, the authors detected DNA-methylation defects that were specific to NT ES cells at some small genomic regions. Most instances of abnormal methylation in NT ES cells were attributable to incomplete reprogramming, leading to epigenetic traces of the mature cell being retained. By contrast, most methylation aberrations found in iPS cells arose owing to abnormal addition of methyl groups during the course of reprogramming. Clearly, neither reprogramming method is perfect.

To build on these intriguing insights, future studies must analyse the reprogramming of mature cell types in a larger cohort of donors. This will be necessary to better understand to what extent epigenetic variance is brought about by differing reprogramming techniques. Testing alternative transcription-factor cocktails may also improve iPS cell quality—it has been found that the combinations or relative amounts of transcription factors used during the generation of mouse iPS cells influence the cells' epigenetic properties¹⁰.

Importantly, one possible explanation for the improved reprogramming outcome observed with SCNT is that the protocol routinely involves using inhibitors of the enzyme histone deacetylase, which facilitate reprogramming by preventing the removal of certain epigenetic marks and so promoting chromosome decompaction. Finally, it will be interesting to test whether recently devised¹¹ non-conventional cell-growth conditions (which alter the epigenetic state of human pluripotent cells, leading them to adopt a more naive state) might alleviate certain defects.

In summary, there is an abundance of factors that can be used to reprogram cells and expand them *in vitro*, and each can influence the epigenetic and functional properties of reprogrammed cells in distinct ways. This complexity disrupts simplistic attempts to define and obtain 'perfect' stem cells. Ma and colleagues' valuable findings may prompt scientists and clinicians to redefine their quest for perfection as a hunt for the adequate, cost-effective and safe. ■

Vladislav Krupalnik and Jacob H. Hanna are in the Department of Molecular Genetics, The Weizmann Institute of Science, Rehovot 76100, Israel.
e-mail: jacob.hanna@weizmann.ac.il

1. Tachibana, M. *et al.* *Cell* **153**, 1228–1238 (2013).

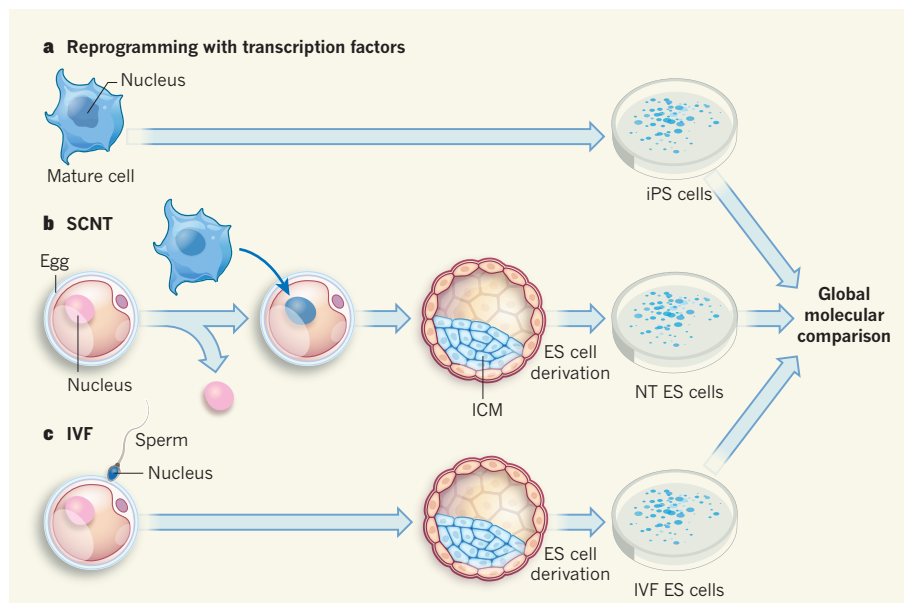


Figure 1 | Comparing techniques for generating stem cells. Three techniques can be used to generate pluripotent stem cells *in vitro*. **a**, Induced pluripotent stem (iPS) cells are generated from mature cells, which can be directly converted by the addition of a transcription-factor cocktail. **b**, In somatic cell nuclear transfer (SCNT), the nucleus is removed from an egg and replaced with the nucleus from a mature donor cell. As this hybrid cell develops into an embryo, pluripotent stem cells called nuclear transfer embryonic stem (NT ES) cells can be extracted from a region called the inner cell mass (ICM). **c**, Embryos derived from *in vitro* fertilization (IVF) give rise to IVF ES cells that can be extracted from the ICM. Ma *et al.*³ analysed the global molecular differences between cells that had undergone SCNT or iPS cell reprogramming, comparing them to the unprogrammed IVF ES cells.

2. Takahashi, K. & Yamanaka, S. *Cell* **126**, 663–676 (2006).
3. Ma, H. *et al. Nature* **511**, 177–183 (2014).
4. Dimos, J. T. *et al. Science* **321**, 1218–1221 (2008).
5. Kim, K. *et al. Nature* **467**, 285–290 (2010).

6. Yamada, M. *et al. Nature* <http://dx.doi.org/10.1038/nature13287> (2014).
7. Thomson, J. A. *et al. Science* **282**, 1145–1147 (1998).
8. Gore, A. *et al. Nature* **471**, 63–67 (2011).

9. Bock, C. *et al. Cell* **144**, 439–452 (2011).
10. Carey, B. W. *et al. Cell Stem Cell* **9**, 588–598 (2011).
11. Gafni, O. *et al. Nature* **504**, 282–286 (2013).

This article was published online on 2 July 2014.

NEUROSCIENCE

A structure to remember

NMDA receptors are crucial in the workings of the brain and in its disorders. Two structures of almost complete receptors reveal the intricate complexity of these large, multi-domain molecular machines. SEE ARTICLE P.191

DAVID STROEBEL & PIERRE PAOLETTI

Since their discovery more than 30 years ago^{1–3}, *N*-methyl-D-aspartate receptors (NMDARs) have fascinated neuroscientists. They are key mediators of synaptic plasticity, the cellular mechanism mediating information storage in the brain. Moreover, dysfunctions of NMDARs are implicated in various neuropsychiatric disorders, from schizophrenia to mental retardation and epilepsy, making them targets of therapeutic interest⁴. Two studies — one from Karakas and Furukawa⁵ published in *Science*, and another by Lee and colleagues⁶ on page 191 of this issue — present the long-awaited molecular structure of an NMDAR subtype, the GluN1–GluN2B receptor, from rats and

frogs, respectively. These atomic maps offer unprecedented views and a wealth of information about a key brain receptor.

Together with the AMPA receptors (AMPA) and kainate receptors, the NMDARs belong to the superfamily of ionotropic glutamate receptors (iGluRs), which mediate excitatory communication between neurons in the central nervous system. The iGluRs are massive protein complexes embedded in the neuronal cell membrane. Each complex contains more than 3,400 amino-acid residues, and is composed of four subunits. Each subunit has a typical modular architecture made up of a large extracellular amino-terminal domain (ATD) that participates in subtype-specific receptor assembly and modulation; a ligand-binding domain (LBD) that

binds receptor-activating agonist molecules; a transmembrane domain (TMD) that forms an ion-channel pore; and a cytoplasmic carboxy-terminal domain (CTD) involved in receptor trafficking and coupling to intracellular signalling molecules.

However, NMDARs differ from other iGluRs in several respects. First, their transmembrane pore is highly permeable to calcium ions and is blocked by magnesium ions — features that are essential for triggering synaptic plasticity. Second, their LBDs require two different agonists for activation. And finally, their ATDs form a major regulatory region that harbours several binding sites for subunit-specific modulators⁴.

In a major feat, the first structure of a full-length iGluR was solved in 2009 — that of the AMPA GluA2 receptor⁷, which has four identical subunits. By contrast, structures of only bits of NMDARs have been solved^{8,9}, and a complete structure showing ATDs linked to the gating core region (which forms from the LBDs and the TMD) has eluded all efforts. Solving the X-ray structure of an NMDAR is particularly challenging because it is formed from two types of subunit (two GluN1 and two GluN2 subunits) and is predicted to have a highly mobile ATD^{10–12}.

Both Lee *et al.* and Karakas and Furukawa raced to solve the structure of GluN1–GluN2B using an approach similar to that used to solve the GluA2 receptor. This involved introducing a flurry of mutations at strategic locations in the receptor to trim flexible regions, including the entire CTD and some loops; reduce surface entropy; remove potential sites of sugar attachment and protein degradation; and eliminate reactive cysteine amino-acid residues, to prevent them from crosslinking. Each group got stuck because their protein crystals diffracted X-rays only to low resolution, presumably because of the flexibility of the molecules. To remove this barrier, both groups crosslinked the ATDs using disulphide bonds, restricting ATD conformational mobility. This improved the resolution of their structures to a reasonable 3.7–3.9 Å, but reduced the channel's ability to function quite drastically. The two teams eventually converged on remarkably similar structures, which probably represent a receptor trapped in an inactive or 'desensitized-like' state.

The GluN1–GluN2B receptor resembles a hot-air balloon, with the TMD at the 'bottom', the ATDs at the 'top' and the LBDs sandwiched in between (Fig. 1) — reminiscent of the layered organization of the GluA2 AMPAR. Notably, the GluN1 and GluN2B subunits alternate around the central pore, settling the lingering debate about subunit order. Also like

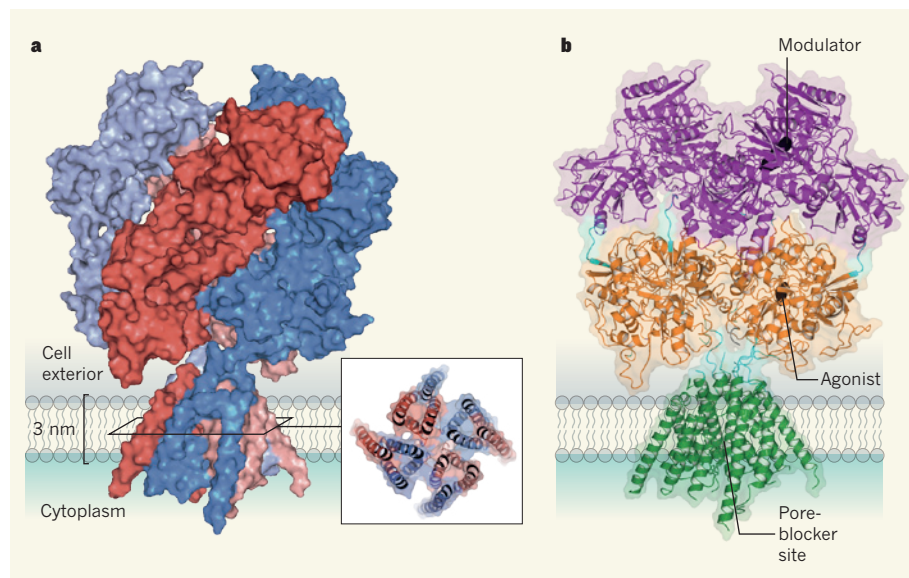


Figure 1 | Structure of an NMDA receptor. Karakas and Furukawa⁵ and Lee *et al.*⁶ independently report the first X-ray structures of an NMDA receptor — the GluN1–GluN2B receptor. The large cytoplasmic domain is absent from both structures. This figure is based on Lee and colleagues' structure. **a**, In this surface representation of the receptor, the two GluN1 subunits are in different shades of blue, and the two GluN2B subunits are in different shades of red. In the membrane region, the GluN1 and GluN2B subunits alternate around a closed central pore (inset). **b**, The receptor displays a layered architecture with a transmembrane domain (TMD; green) that forms an ion channel, four ligand-binding domains (LBDs; orange) and four amino-terminal domains (ATDs; purple). Short linking segments (cyan) connect the ATDs to the LBDs, and the LBDs to the TMD. A modulator and agonist are shown in black. The binding site for pore blockers is indicated.

GluA2, the four subunits assemble as a dimer-of-dimers and display 'domain swapping', such that pairing of domains between the ATD and LBD layers involves different subunits. Domain swapping thus seems to be a distinct, evolutionarily conserved structural feature of iGluRs that ensures subunit intertwining and overall cohesion of the large extracellular region.

The conservation of the tetrameric organization between NMDARs and AMPARs is somewhat surprising, considering the major differences in their ATD structures and in the ATD connections to the gating core. In the GluN1–GluN2B structure, the ATD heterodimers adopt the same peculiar twisted structure previously observed⁹ for isolated ATD domains. Moreover, the ATDs pack tightly against the LBDs, whereas ATD–LBD interactions in the GluA2 AMPAR are minimal. The extracellular region is therefore much more compact in NMDARs than in AMPARs; this characteristic is probably related to the unique role of NMDAR ATDs in modulating channel activity⁴.

Lee *et al.* provide the first structure of an almost complete iGluR pore, including the crucial pore loop — the short region that controls ion permeation. As expected¹³, the pore resembles an inverted potassium channel, and the symmetry of the pore loop indicates different contributions of the GluN1 and GluN2 subunits to permeation, as was previously predicted by electrophysiological studies¹⁴. Although neither research group provides a picture of a magnesium ion in its blocking site, Lee *et al.* do provide the first glimpse

of channel-blocker sites, and Karakas and Furukawa highlight a potential calcium-ion storage site at the entrance of the pore. But more work is needed, including the determination of higher-resolution structures, to get a deeper understanding of the singular permeation properties of NMDARs.

What is the physiological relevance of the state captured in the structures, and what do the structures tell us about functional coupling between the ATD and gating-core regions? By assaying for domain proximity, Karakas and Furukawa convincingly show that the distinct inter-domain and inter-subunit interfaces observed in their crystal structure are present in an intact membrane-embedded receptor. When electrophysiological recordings of the crystallized receptors were made, however, the receptors were found to be functionally 'silent'. Functionality was recovered only when the authors broke the ATD crosslinks using reducing agents. This suggests that the conformational state of the ATDs in a functional receptor differs from that in the crystal. In a related result, Lee *et al.* observed that, in their low-resolution structure, the ATDs adopt a conformation in which the two ATD dimers split apart.

How conformational changes of the ATDs modulate receptors by influencing the gating core remains to be decrypted. In particular, the proposed^{10–12} large-scale motions of individual ATDs will have to be reconciled with the presence of extensive, presumably motion-restricting, interactions between the lower lobes of the ATDs and the upper lobes of the

LBDs. Understanding the structural plasticity of the multi-domain receptor and the functional coupling between the various layers remains a major challenge. Nevertheless, the two new structures offer an exceptional framework for future structural, pharmacological and functional studies of NMDARs and of neurotransmission in general. ■

David Stroebel and Pierre Paoletti are at the *Institute of Biology, École Normale Supérieure, CNRS UMR8197, INSERM U1024, 75005 Paris, France.*

e-mails: pierre.paoletti@ens.fr; david.stroebel@ens.fr

1. Watkins, J. C. & Jane, D. E. *Br. J. Pharmacol.* **147**, S100–S108 (2006).
2. Nowak, L., Bregestovski, O., Ascher, P., Herbet, A. & Prochiantz, A. *Nature* **307**, 462–465 (1984).
3. Mayer, M. L., Westbrook, G. L. & Guthrie, P. B. *Nature* **309**, 261–262 (1984).
4. Paoletti, P., Bellone, C. & Zhou, Q. *Nature Rev. Neurosci.* **14**, 383–400 (2013).
5. Karakas, E. & Furukawa, H. *Science* **344**, 992–997 (2014).
6. Lee, C.-H. *et al. Nature* **511**, 191–197 (2014).
7. Sobolevsky, A. I., Rosconi, M. P. & Gouaux, E. *Nature* **462**, 745–756 (2009).
8. Furukawa, H., Singh, S. K., Mancuso, R. & Gouaux, E. *Nature* **438**, 185–192 (2005).
9. Karakas, E., Simorowski, N. & Furukawa, H. *Nature* **475**, 249–253 (2011).
10. Zhu, S., Stroebel, D., Yao, C. A., Taly, A. & Paoletti, P. *Nature Struct. Mol. Biol.* **20**, 477–485 (2013).
11. Dutta, A., Shrivastava, I. H., Sukumaran, M., Greger, I. H. & Bahar, I. *Structure* **20**, 1838–1849 (2012).
12. Burger, P. B. *et al. Mol. Pharmacol.* **82**, 344–359 (2012).
13. Chen, G.-Q., Cui, C., Mayer, M. L. & Gouaux, E. *Nature* **402**, 817–821 (1999).
14. Wollmuth, L. P., Kuner, T. & Sakmann, B. *J. Physiol. (Lond.)* **506**, 13–32 (1998).

CLIMATE SCIENCE

City heat

Temperatures in cities are often higher than in surrounding rural areas, an effect commonly thought to be associated with lower evaporative cooling from urban land. But on page 216 of this issue, Zhao *et al.* report that the phenomenon is caused by variations in heat convection to the lower atmosphere (L. Zhao *et al. Nature* **511**, 216–219; 2014).

The authors analysed temperature data for 65 North American cities using computational models. They found that, in humid regions, convection dissipates heat less efficiently from cities than over rural land, because there is greater aerodynamic resistance to heat diffusion. The opposite is true in dry regions.

Furthermore, the temperature difference between cities and rural areas in humid regions is even greater in dry years, probably because of feedback between precipitation and temperature. The researchers suggest that reducing anthropogenic heat won't ease city-dwellers' discomfort, but increasing urban albedo — heat reflectivity — will. **Andrew Mitchinson**



OCEAN CHEMISTRY

Fingerprints of a trace nutrient

Lack of dissolved iron in the sea limits biological productivity and the uptake of carbon dioxide. The sources of dissolved iron in the North Atlantic Ocean have been identified from isotopic variations of this trace nutrient. [SEE LETTER P.212](#)

JOSEPH A. RESING & PAMELA M. BARRETT

Iron constitutes 5% of Earth's crust, yet the element is sparse in the vast expanse of the global ocean — every litre of surface sea water contains, on average, less than 20 nanograms of iron¹. Despite this, iron has a crucial role in regulating primary production in the oceans (the synthesis of organic matter from carbon dioxide by phytoplankton). On page 212 of this issue, Conway and John² examine the relative importance of different sources of iron to the North Atlantic Ocean. Constraining the magnitude of the contributions from the various iron sources will allow us to incorporate iron cycling more accurately into global climate models and to better understand its role in regulating oceanic carbon export.

The sources of oceanic iron include natural dust from arid lands, anthropogenic

aerosols, rivers, coastal sediments and submarine volcanoes. In the past decade, new analytical techniques have made it possible to look at variations in two natural isotopes of iron (expressed as $\delta^{56}\text{Fe}$) at the concentrations found in sea water, providing a crucial tool with which to understand how iron moves through the oceans. Conway and John have produced the first $\delta^{56}\text{Fe}$ section across an ocean basin as a part of the international GEOTRACES programme, which aims to understand trace-element distribution and cycling in the oceans.

The authors' $\delta^{56}\text{Fe}$ data allow them to 'fingerprint' dissolved iron from different sources and apportion each source's contribution to the pool of this nutrient in the North Atlantic. They find that the largest source is wind-borne Saharan dust, followed by the release of iron from ocean sediments along coastal margins. The smallest but most

steady source is from submarine volcanic and hydrothermal emissions, which have continuously released iron to the oceans over millions of years.

Multiple feedbacks act between iron delivery to the oceans and global climate. Phytoplankton, the base of the oceanic food web, requires iron as an essential trace nutrient for diverse biochemical pathways that control processes such as photosynthesis and nitrogen fixation. When phytoplankton sink into the deep sea, they take with them any carbon they have accumulated. This represents one mechanism by which oceans take up CO_2 from the atmosphere and transport it to the deep ocean, thereby modulating CO_2 levels in the atmosphere. Expansive regions of the world ocean, including the subarctic Pacific, eastern equatorial Pacific and Southern oceans, are rich in macronutrients such as nitrate, phosphate and silicate, but have relatively low biological productivity because of the extremely low levels of dissolved iron, which limits phytoplankton growth³.

Variations in the supply of iron to the ocean surface have probably contributed to changes in atmospheric CO_2 levels over glacial and interglacial cycles^{3,4}. During glacial times, low sea levels expose iron-rich continental shelves to erosion, and high aridity and wind speeds increase the area of dry regions and the amount of dust that becomes wind-borne from these areas⁵. Ice and marine-sediment records, together with modelling results, all indicate that iron fluxes to the ocean were 2–20 times higher during glacial periods than during the current interglacial, enhancing productivity and carbon export during those periods⁶.

Superimposed on this variability is the steady supply of iron from hydrothermal vents at submarine volcanoes that circle the globe in both the shallow and deep ocean. There is evidence⁷ that iron from hydrothermal vents can become organically bound or stabilized as colloids (extremely small particles suspended in water), allowing it to persist over great length and time scales. Iron from deep hydrothermal sources might ultimately reach the nutrient-rich surface waters of the Southern Ocean, where iron abundance regulates phytoplankton growth and thus CO_2 uptake⁸. Shallow volcanoes in the western Pacific may also supply iron to the equatorial undercurrent, providing a pathway for iron into the ocean interior and the nutrient-rich eastern equatorial Pacific.

Differences in the magnitude of iron sources used in climate models produce wide variations in the amount of CO_2 taken up by the oceans during glacial periods as a result of iron fertilization^{5,6,9–11}. Data on iron isotopes will better constrain oceanic iron budgets, improving model parameterizations and thus allowing a better understanding of the effects of changing iron input on past and future climate.

Many outstanding questions remain about

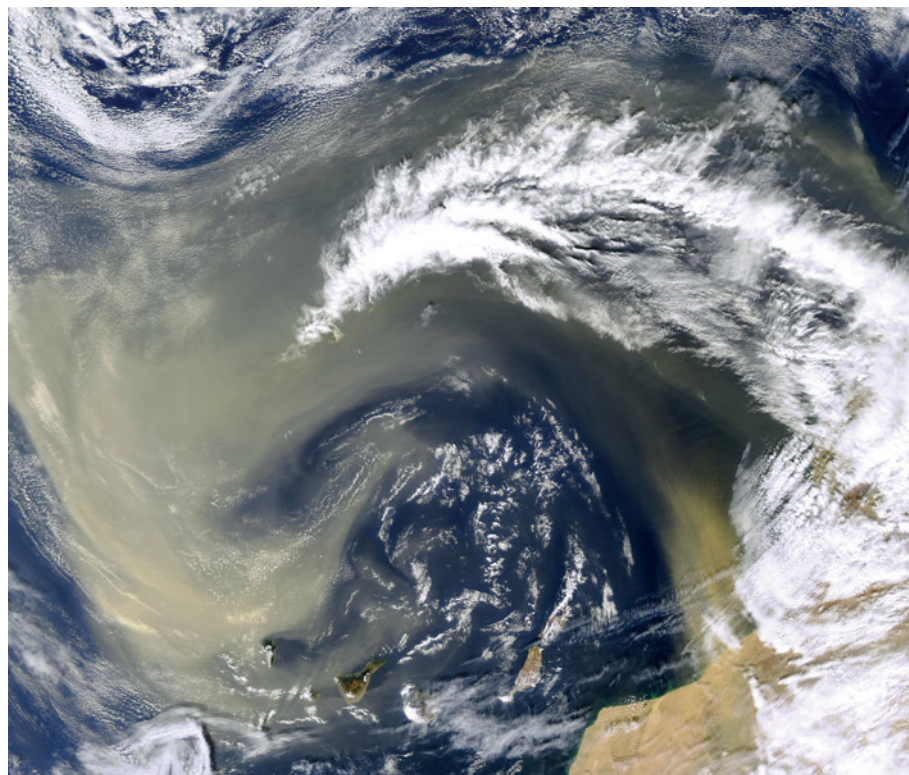


Figure 1 | Saharan dust blows north of the Canary Islands. Conway and John² report that wind-borne Saharan dust is the greatest source of dissolved iron in the North Atlantic Ocean.

JACQUES DESCLOITRES, MODIS RAPID RESPONSE TEAM, NASA/GSFC

the oceanic processes that fractionate iron isotopes. For example, Conway and John attribute the major fraction of dissolved iron in their samples to natural aerosol dust, but the isotope ratios of this fraction are unexpectedly 'heavy' — there is a high abundance of heavier isotopes — given that aerosols derived from arid lands have lighter isotopic signatures. The authors argue that dissolution and complexation processes produce the heavy $\delta^{56}\text{Fe}$ value that they use to identify aerosol-dust iron sources. It is easy to accept this hypothesis, because natural aerosol dust is the obvious major source of iron to the North Atlantic (Fig. 1). However, in regions in which natural dust is a less dominant iron source, it might be more difficult to attribute $\delta^{56}\text{Fe}$ signatures to specific sources. It is therefore crucial to gain a better understanding of how iron fractionates as it cycles between dissolved and particulate phases, and between the surface and the deep ocean.

As measurements of $\delta^{56}\text{Fe}$ in the ocean increase, there is great potential to fingerprint currently uncharacterized sources of iron, and to identify other oceanic processes by their $\delta^{56}\text{Fe}$ signature. For example, iron from anthropogenic aerosols and from natural dust affected by anthropogenic emissions are likely to have distinct iron-isotopic ratios — the combustion of fossil fuels and biomass produce easily solubilized aerosol iron¹² as well as acids that probably enhance the solubility of natural aerosols¹³. This effect may be especially pronounced in the North Pacific, for which increased industrialization and coal burning in Asia may cause a shift in the flux

of bioavailable iron, with unknown implications for the $\delta^{56}\text{Fe}$ signature of dissolved iron. The use of isotopes to map the distribution of iron sources to the global ocean is still young, but with so much still to discover, holds great promise. ■

Joseph A. Resing is at the Joint Institute for the Study of the Atmosphere and Ocean, University of Washington, and NOAA-PMEL, Seattle, Washington 98115 USA. **Pamela M. Barrett** is in the Department of Oceanography, University of Washington, Seattle, Washington 98195 USA.
e-mails: resing@uw.edu; barrettp@uw.edu

1. Johnson, K. S., Gordon, R. M. & Coale, K. H. *Mar. Chem.* **57**, 137–161 (1997).
2. Conway, T. M. & John, S. G. *Nature* **511**, 212–215 (2014).
3. Boyd, P. W. & Ellwood, M. J. *Nature Geosci.* **3**, 675–682 (2010).
4. Martin, J. H. *Paleoceanography* **5**, 1–13 (1990).
5. Mahowald, N. M. et al. *Glob. Biogeochem. Cycles* **19**, GB4030 (2005).
6. Kohfeld, K. & Ridgwell, A. in *Surface Ocean–Lower Atmosphere Processes* (eds Le Quééré, C. & Saltzman, E. S.) 251–286 (AGU, 2009).
7. Bennett, S. A. et al. *Earth Planet. Sci. Lett.* **270**, 157–167 (2008).
8. Tagliabue, A. et al. *Nature Geosci.* **3**, 252–256 (2010).
9. Ziegler, M., Diz, P., Hall, I. R. & Zahn, R. *Nature Geosci.* **6**, 457–461 (2013).
10. Martínez-García, A. et al. *Nature* **476**, 312–315 (2011).
11. Tagliabue, A., Aumont, O. & Bopp, L. *Geophys. Res. Lett.* **41**, 920–926 (2014).
12. Sedwick, P. N., Sholkovitz, E. R. & Church, T. M. *Geochem. Geophys. Geosyst.* **8**, Q10Q06 (2007).
13. Baker, A. R. & Croot, P. L. *Mar. Chem.* **120**, 4–13 (2010).

This article was published online on 2 July 2014.



50 Years Ago

Dr. J. B. Mays has in recent years made a significant contribution towards our understanding of some of the intractable problems of juvenile delinquency, notably in his investigation, published in 1959 ... The central thesis of this investigation is that “society itself is in some respects criminogenic and to a considerable extent precipitates its own delinquency” ... but what is unusual is the extent to which Dr. Mays tries to prove that “criminal behaviour arises naturally enough out of the ordinary daily affairs in which all of us are engaged” ... Dr. Mays has had to rely in the main on his own experiences as a youth club leader, and on the research he carried out among the boys of the youth club with which he was connected ... The interesting question is, would these same boys have committed offences had they been born and brought up in any other part of Britain?

From *Nature* 11 July 1964

100 Years Ago

We have received a copy of a report on the progress of the metric system which was presented by Dr. Guillaume at the meeting of the fifth general conference on weights and measures held in Paris in October last ... A section is devoted to legislation with respect to the metric system in various countries since the fourth general conference. During the past six years the system has been made obligatory in several countries, notably Denmark, Siam, the Belgian Congo, and certain of the republics of Central America. Dr. Guillaume considers that the difficulties standing in the way of the adoption of the system in Great Britain and the United States have been greatly exaggerated by its opponents.

From *Nature* 9 July 1914

BEHAVIOURAL ECONOMICS

A caring majority secures the future

An experiment studying people's willingness to sacrifice personal gains so that resources are passed to future generations shows that this occurs only when extractions by free-riders are curbed by majority rule. [SEE LETTER P.220](#)

LOUIS PUTTERMAN

What kind of world will we pass on to those who follow us? Most of us would like to leave a liveable, green planet, but are ‘most of us’ enough to turn current environmental trends around, and are we willing to sacrifice enough to make a real difference? In this issue, Hauser et al.¹ (page 220) present the results of a decision-making experiment, involving participants hired through an online job marketplace, that helps us to think more clearly about these questions. The authors found impressive evidence

of majority willingness to sacrifice personal gains for future generations, but they show that whether such majorities are adequate for the task depends on whether choices are made individually or by group decision.

Social dilemmas are those all-too-common situations in which an individual's pursuit of self-interest leads to a collectively inferior outcome. The common-pool resource game (with its familiar ‘tragedy of the commons’ prediction) and the prisoner's dilemma game are two well-known examples that can demonstrate such outcomes². Public-choice theory suggests that optimal decisions are possible in social

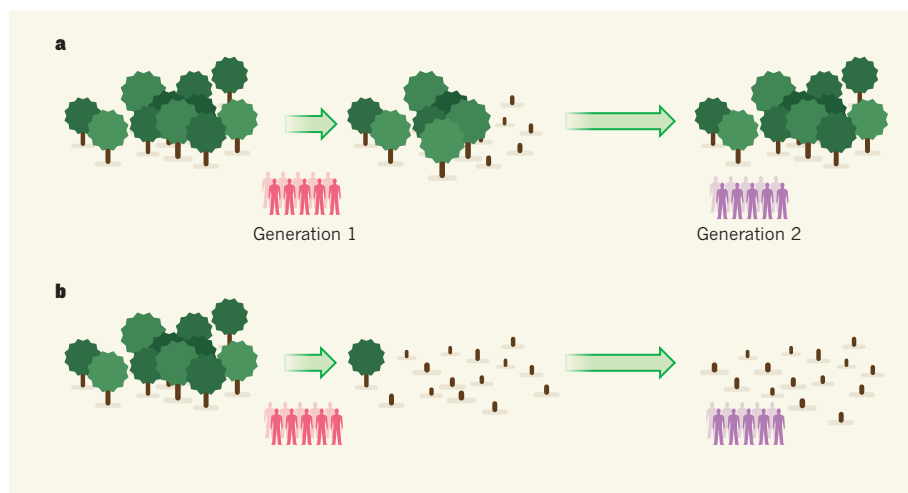


Figure 1 | Leaving enough for those to come. **a**, In Hauser and colleagues' experiment¹, each generation can pass a self-replenishing pool of resources on to the next generation, but only if over-extraction from that pool is avoided. The experiment is analogous to the living human population's stewardship of Earth's biological, geological and atmospheric resources. **b**, Over-extraction, in the experiment, means the next generation gets nothing.

dilemmas if groups can vote on rules that will be binding on all participants³. However, the dilemma game that Hauser *et al.* studied does not automatically yield to voting because it is an intergenerational game, so most of the affected parties represent unborn generations and thus cannot be in the present voting group.

The authors' game involves groups (generations) of five individuals, who must decide individually or collectively how much to withdraw for themselves from a common resource pool that can replenish itself and be passed along to a next generation only if the sum of extractions is below a critical level (Fig. 1). Many such generations can potentially benefit from the resource pool, but if the members of any one generation think of themselves only, future generations will be left with nothing. Under these circumstances, strong concern for future generations is necessary if an outcome that is jointly optimal for all potential generations is to be achieved.

Traditional game theory assumed that people were not only rational but also strictly selfish. However, social scientists, including economists, have been amassing evidence that many people are at least conditionally willing to act for the greater good if assured that others will also do so⁴. Hauser and colleagues' paper stands at the crossroads of the behavioural study of social preferences, as economists call them, and the formal logic of public-choice theory.

In contrast to the vibrant social-dilemmas field as a whole, the intersection of public-choice analysis with behavioural economics has been understudied until recently. Social-dilemma experiments show that cooperation tends to diminish over time if the cooperatively inclined have no way of defending themselves against self-serving 'free riders'. Empowering cooperators with the ability to punish free-riders helps in some settings⁵, but unless

appropriate norms or enforcement mechanisms are worked out, benefits may be undermined by retaliation and feuding⁶. Enforcing rules that have been agreed by a vote is a mechanism that can work well if conditional cooperativeness is common and if antisocial or strictly selfish propensities mark only a minority⁷.

In their intergenerational game, Hauser and colleagues find that, when decisions are made by individuals acting independently, the resource pool usually collapses despite majority pro-sociality, because it takes only one or two over-extractors to push the combined withdrawal of resources past the point of no return. When decisions are made by majority voting, however, these over-extractors are outnumbered. And, in an important bonus, the majority's ranks actually swell thanks to the assurance that voting gives to conditional cooperators that they will not be taken advantage of.

The degree of concern that Hauser and colleagues' participants exhibited for members of later groups, whom they will never know, is likely to surprise many. These individuals, who were hired from the online marketplace Amazon Mechanical Turk and are presumably relying on this part-time work to make ends meet, had to voluntarily forgo half of their prospective bonus to pass the resource along to the next group. Although similar degrees of concern for fairness have been exhibited by participants in other experimental studies⁸, it must be a lot easier for individuals to sacrifice a bonus of the order of pocket change than for societies to vote for major sacrifices, such as trading in their cars for bicycles and public transport. In other ways, however, the authors' experimental design favours less, rather than more, sacrifice: most real-world individuals have direct concerns for their children and

grandchildren, which confer an emotional stake in the future, but there are no analogous connections for the anonymous online employees.

There are also other difficulties with extrapolating this experiment's remarkable results to the solution of global environmental issues. One major problem lies in the fact that the enforcement machinery that would be needed to put Hauser and colleagues' mechanism to work on a global scale does not yet exist — as exemplified by the lack of international consensus over the Kyoto Protocol. Thus, the encouraging finding that majorities are willing to sacrifice if they are assured that others will join in, and the recognition that majority rule can constrain selfish minorities, must be matched by sufficient assent to adopting the relevant enforcement structure.

On this point, existing evidence offers a glimmer of hope. Recent decision experiments^{9,10} let participants determine, by voting, whether to adopt rules that were made binding on group members by administrative punishment, or to operate in a regime with only peer-to-peer sanctioning available. When the centralized-enforcement regime was available at low cost, these studies found that majorities tended to favour it. Members of the minority that had displayed the antisocial behaviour of punishing cooperators often voted against the regime, but they were outnumbered. Hauser and colleagues' findings provide further evidence that a behavioural approach to public choice can help to elucidate issues of first-order importance, and that experimental techniques can complement theoretical analysis and observations in natural settings to yield insights that may prove to be crucial to a sustainable society. ■

Louis Putterman is in the Department of Economics, Brown University, Providence, Rhode Island 02912, USA.

e-mail: louis_putterman@brown.edu

1. Hauser, O. P., Rand, D. G., Peysakhovich, A. & Nowak, M. A. *Nature* **511**, 220–223 (2014).
2. Ostrom, E., Gardner, R. & Walker, J. *Rules, Games, and Common-Pool Resources* (Univ. Michigan Press, 1994).
3. Feldman, A. M. & Serrano, R. *Welfare Economics and Social Choice Theory* (Springer, 2006).
4. Fischbacher, U. & Gächter, S. *Am. Econ. Rev.* **100**, 541–556 (2010).
5. Van Lange, P., Rockenbach, B. & Yamagishi, T. (eds) *Reward and Punishment in Social Dilemmas* (Oxford Univ. Press, 2014).
6. Nikiforakis, N. & Engelmann, D. *J. Econ. Behav. Organ.* **78**, 319–332 (2011).
7. Ertan, A., Page, T. & Putterman, L. *Eur. Econ. Rev.* **53**, 495–511 (2009).
8. Cooper, D. & Kagel, J. in *The Handbook of Experimental Economics* Vol. 2 (eds Kagel, J. & Roth, A.) (Princeton Univ. Press, in the press).
9. Kamei, K., Putterman, L. & Tyran, J.-R. *Exp. Econ.* <http://dx.doi.org/10.1007/s10683-014-9405-0> (2014).
10. Markussen, T., Putterman, L. & Tyran, J. R. *Rev. Econ. Stud.* **81**, 301–324 (2014).

This article was published online on 25 June 2014.

Metabolism of stromal and immune cells in health and disease

Bart Ghesquière^{1,2}, Brian W. Wong^{1,2}, Anna Kuchnio^{1,2} & Peter Carmeliet^{1,2}

Cancer cells have been at the centre of cell metabolism research, but the metabolism of stromal and immune cells has received less attention. Nonetheless, these cells influence the progression of malignant, inflammatory and metabolic disorders. Here we discuss the metabolic adaptations of stromal and immune cells in health and disease, and highlight how metabolism determines their differentiation and function.

S Stromal cells are connective cells that support parenchymal cells. Aside from endothelial cells (ECs), which build blood vessels to supply oxygen and nutrients^{1,2}, other stromal cells include fibroblasts, which maintain the structural framework in tissues. Another class of cells that maintains tissue homeostasis is immune cells. The immune system collectively functions to protect tissues from infections and foreign antigens. Notably, several of these cell types not only contribute to disease progression but also to therapy resistance³.

Cancer cells undergo numerous metabolic adaptations; a result of oncogenic transformation, gene mutations or altered expression of metabolic enzymes^{4–7}. Metabolism of stromal and immune cells has received less attention, yet they represent an abundant cell population in tissues and tumours, even up to 80–90% in some desmoplastic tumours⁸. We focus on two types of immune cells (T cells and macrophages) and two types of stromal cells (ECs and fibroblasts), for which metabolism has been best characterized. Emerging evidence indicates that metabolism not only co-determines their differentiation and function, but also that metabolic changes in these cells contribute to the pathogenesis of cancer, diabetes and inflammatory diseases.

Vascular endothelial cells

ECs are quiescent for years, but when tissues are deprived of oxygen or nutrients, they sprout to vascularize tissues. Vessel sprouting is a highly coordinated process, relying on a migrating endothelial ‘tip’ cell at the forefront that guides the vessel sprout (but rarely proliferates), and on trailing endothelial ‘stalk’ cells elongating the sprout via proliferation^{1,9}. Vascular endothelial growth factor (VEGF) is a tip cell signal, whereas Delta-like 4 (DLL4)/Notch signalling is a stalk cell signal. Tip versus stalk cell phenotypes are not predetermined cell fates, but are interchangeable. By competing for the tip position, ECs continuously secure optimal fitness to lead the sprout. Although many genetic signals regulate this process, the role of EC metabolism in vessel sprouting has only recently been considered. We postulated that this angiogenic switch requires a metabolic switch¹⁰.

Endothelial cells are glycolytic

ECs have immediate access to oxygen in the blood, but do not take advantage of its availability to adopt an oxidative metabolism. Instead, oxygen consumption is lower and the mitochondrial volume fraction is smaller than in oxidative cell types¹¹. Also, inhibition of mitochondrial ATP production or stimulation of oxygen consumption does not affect vessel sprouting¹⁰. This is in line with reports that EC mitochondria serve a signalling function via the production of pro-angiogenic reactive oxygen species (ROS) rather than act as a bioenergetic power source¹². Instead, ECs are highly glycolytic, more so than other cell types, and generate up to 85% of

their ATP via glycolysis¹⁰ (Fig. 1a). In fact, their glycolytic rates are comparable to those in many cancer cells, known to have a high aerobic glycolytic metabolism¹⁰. ECs are glucose addicted, as glucose starvation or treatment with the glucose analogue 2-deoxy-D-glucose induces cytotoxicity¹³. Although quiescent ECs require substantial glycolysis for homeostatic maintenance, glycolysis is still increased (by twofold) when shifting to proliferation¹⁰.

A recent report studied the role of glycolysis in vessel sprouting by focusing on 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase-3 (PFKFB3)¹⁰. Its kinase activity is 700-fold higher than its bisphosphatase activity and favours the production of fructose-2,6-bisphosphate (F2,6BP), an allosteric activator of phosphofructokinase-1 (PFK1), itself a rate-limiting enzyme of glycolysis (Fig. 1a). Inhibition of PFKFB3 reduces EC glycolytic flux levels, but by no more than 30–40%. Nonetheless, this is sufficient to impair vessel sprouting *in vitro* and induce vascular hypobranched *in vivo*¹⁰. Mechanistically, PFKFB3 deficiency in ECs reduces proliferation of stalk cells, in agreement with reports that PFKFB3-driven glycolysis is elevated in the G₁ cell cycle phase in preparation for the S phase. PFKFB3-driven glycolysis is also necessary for tip cell function (that is, migration) and competitiveness by providing energy for the remodelling of the cytoskeleton. Indeed, when ECs form lamellipodia to migrate, glycolytic enzymes are relocated to these protrusions, where they become associated with F-actin. This compartmentalization enhances local synthesis of ATP at the leading membrane ruffles in lamellipodia¹⁰.

Treatment of mice with a PFKFB3 blocker reduces glycolysis only partially and transiently, yet impairs pathological angiogenesis, indicating that ECs are sensitive to small changes in glycolysis¹⁴. More complete and sustained inhibition of glycolysis by 2-deoxy-D-glucose impairs angiogenesis but is more toxic. Hence, partial and transient blockade of glycolysis may provide a new paradigm for anti-angiogenic therapy.

There are several hypotheses attempting to explain why ECs prefer glycolysis even though sufficient oxygen is present for oxidative metabolism (‘aerobic glycolysis’) and despite the fact that glucose oxidation is more efficient in yielding ATP (only 2 moles of ATP are generated via glycolysis versus up to 36 moles of ATP via glucose oxidation, per mole of glucose). First, by minimizing oxygen consumption, quiescent ECs reserve more oxygen for transfer to perivascular tissue cells. Second, ECs must vascularize avascular tissues, where low oxygen levels would preclude oxidative metabolism. Third, because of its high rate, glycolysis can generate more ATP than oxidative phosphorylation (OxPhos) as long as glucose availability is not limiting¹⁵. In fact, since interstitial oxygen levels drop faster than glucose levels over a distance away from a blood vessel, ECs

¹Laboratory of Angiogenesis and Neurovascular Link, Vesalius Research Center, Department of Oncology, University of Leuven, Leuven B-3000, Belgium. ²Laboratory of Angiogenesis and Neurovascular Link, Vesalius Research Center, VIB, Leuven B-3000, Belgium.

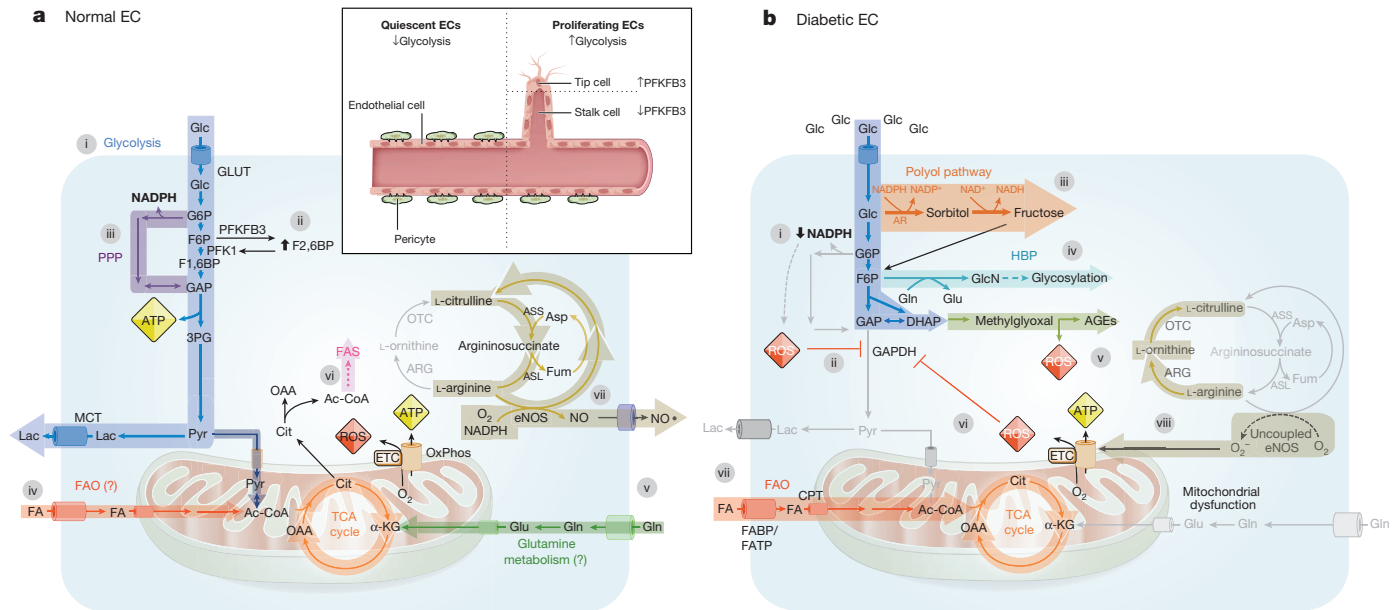


Figure 1 | Endothelial cell metabolism. For all figures: important metabolic pathways onto which cells rely are shown in colour; less critical or studied pathways are shown in grey. **a**, Scheme of the metabolic pathways in a normal EC. (i) ECs use glucose as primary source for energy production and are highly glycolytic (blue). Only a minor fraction of glucose is oxidized in the TCA cycle (orange). (ii) PFKFB3 is a regulator of glycolysis via conversion of F1,6BP to F2,6BP, an activator of PFK1. (iii) The PPP generates NADPH for redox homeostasis (purple). FAO generates acetyl-CoA (red) (iv), whereas glutamine metabolism generates α -KG (green) (v), which are metabolized in the TCA cycle (clockwise direction; orange). However, their importance versus glycolysis is unclear (denoted by question marks). (vi) ECs also synthesize fatty acids (FAS; pink). (vii) NOS converts arginine to nitric oxide (NO) and citrulline, which is recycled to regenerate arginine (tan). Inset: proliferating ECs have higher glycolysis than quiescent ECs; glycolysis is higher in tip than stalk cells. **b**, Metabolic changes in ECs in diabetes. (i) High glucose levels impair the PPP (tan), thereby lowering NADPH and elevating ROS levels. (ii) ROS reduces glycolysis by inactivating GAPDH. (iii) This increases the polyol pathway flux (dark orange), leading to generation of ROS and toxic AGEs. (iv) The flux of the hexosamine biosynthesis pathway (cyan) is also increased, resulting in alteration of glycosylation. (v) GAP and DHAP are converted to methylglyoxal, leading to production of ROS and AGEs (olive). (vi) Excess mitochondrial ROS generation inhibits GAPDH, leading to accumulation of GAP and DHAP and

can continue to rely on anaerobic glycolysis¹⁶. ECs are also resistant to hypoxia as long as glucose is available, but become sensitive to oxygen-deprivation when glucose is limiting¹⁷. Finally, glycolytic intermediates are used for anabolic generation of macromolecules and for redox control, which are necessary for EC proliferation, migration and survival¹⁸. Aside from their metabolic role, glycolytic metabolites (lactate) can stimulate angiogenesis also as signalling molecules (Box 1).

Role of additional metabolic pathways in ECs

The role of other metabolic pathways in angiogenesis *in vivo* remains less well defined (Fig. 1a). The pentose phosphate pathway (PPP), a side pathway of glycolysis, promotes cell proliferation by producing NADPH and ribose-5-phosphate, which are used for the synthesis of lipids, nucleotides and amino acids (histidine). The PPP also provides protection against oxidative stress by controlling redox homeostasis. Indeed, NADPH is used to convert the oxidized form of glutathione (GSSG) to its reduced form (GSH), a major antioxidant¹⁹.

Another fuel used by proliferating cells is L-glutamine (referred to as glutamine), which has various metabolic fates, for example, oxidation of glutamine-derived α -ketoglutarate (α -KG) for energy production, glutaminolysis followed by pyruvate recycling into the tricarboxylic acid (TCA) cycle for continued TCA cycling, reductive carboxylation for lipid

reducing glycolysis. (vii) High glucose levels increase FAO at the expense of glucose oxidation. (viii) Increased arginase levels reduce the availability of arginine for the production of NO by NOS, which is uncoupled/less functional, and produces superoxide that induces mitochondrial dysfunction (tan). 3PG, 3-phosphoglycerate; α -KG, α -ketoglutarate; Ac-CoA, acetyl-coenzyme A; AGEs, advanced glycation end products; AR, aldose reductase; ARG, arginase; ASL, argininosuccinate lyase; Asp, aspartate; ASS, argininosuccinate synthase; Cit, citrate; CPT, carnitine palmitoyltransferase; DHAP, dihydroxyacetone phosphate; eNOS, endothelial nitric oxide synthase; ETC, electron transport chain; F1,6BP, fructose 1,6-bisphosphate; F2,6BP, fructose 2,6 bisphosphate; F6P, fructose 6-phosphate; FA, fatty acid; FABP, fatty acid binding protein; FAO, fatty acid oxidation; FAS, fatty acid synthesis; FATP, fatty acid transfer protein; Fum, fumarate; G6P, glucose 6-phosphate; GAP, glyceraldehyde 3-phosphate; Glc, glucose; GlcN, glucosamine; Gln, glutamine; Glu, glutamate; GLUT, glucose transporter; HBP, hexosamine biosynthesis pathway; Lac, lactate; MCT, monocarboxylate transporter; NADPH, nicotinamide adenine dinucleotide phosphate; NO, nitric oxide; OAA, oxaloacetate; OTC, ornithine transcarbamylase; OxPhos, oxidative phosphorylation; PFK1, phosphofructokinase-1; PFKFB3, 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase-3; PPP, pentose phosphate pathway; Pyr, pyruvate; ROS, reactive oxygen species; TCA, tricarboxylic acid.

synthesis, or polyamine synthesis for mitogenesis (Fig. 1a). In line with findings that blockade of mitochondrial ATP production does not affect vessel sprouting¹⁰, the importance of glutamine oxidation for the production of ATP in ECs is debated^{10,20}. Nonetheless, glutamine metabolism is essential, as inhibition of glutaminase decreases proliferation and induces senescence of ECs^{21,22}. ECs also use glutamine for the production of ornithine, a precursor of polyamines²³. The role of glutaminolysis and reductive carboxylation in relation to vessel sprouting has not been studied to date.

Fatty acids can also serve as fuel for energy production, such as in oxidative cardiomyocytes. In metabolically stressed cancer cells, fatty acid oxidation (FAO) is vital for their survival by generating ATP and ensuring redox homeostasis (Fig. 1a)²⁴. ECs also metabolize fatty acids for ATP production, but like for glutamine oxidation, the importance of FAO relative to glycolysis is debated^{20,25}. Another role of FAO in ECs has not been described yet. Notably, proliferative cells also synthesize lipids, partly to form new membranes for cell division, to generate signalling molecules or to modulate signalling²⁶. For instance, cholesterol promotes vessel sprouting by facilitating the formation of membrane lipid rafts in ECs, necessary for membrane localization and signalling of the pro-angiogenic VEGF receptor (VEGFR)-2 (ref. 27). Conversely, apolipoprotein-B lipoproteins, involved in cholesterol transport, impair angiogenesis

BOX 1

Signalling effects of metabolites and metabolic enzymes

Endothelial cells

Lactate, produced by cancer cells, promotes tumour angiogenesis via mechanisms independent of its role in glycolysis. By inhibiting the oxygen-sensing prolyl hydroxylase 2 (PHD2), lactate activates the hypoxia-inducible transcription factor HIF-1 α in cancer cells and upregulates angiogenic signals⁹⁷. Via the same mechanism, lactate triggers I κ B α degradation, thereby stimulating an autocrine pro-angiogenic NF- κ B/IL-8 pathway⁹⁸. Furthermore, lactate activates the receptor tyrosine kinases AXL, TIE2 and VEGFR-2 in a ligand-independent manner⁹⁹. In agreement, perturbing lactate influx in ECs by blocking the monocarboxylate transporter-1 (MCT1) inhibits tumour angiogenesis²¹. Interestingly, metabolic enzymes may also affect post-translational modification of targets. For instance, EC loss of fatty acid synthase (FASN) impedes angiogenesis by reducing palmitoylation of VEGFR-2 and eNOS, which impairs their plasma membrane targeting and activation¹⁰⁰.

Immune cells

Several metabolites have a signalling role in immune cells. For instance, high levels of the adenylate AMP in nutrient-poor conditions activate AMPK and control a metabolic checkpoint in T cells by acting as a brake on immune cell expansion when energy supplies are poor⁵⁵. Another example is α -KG, which is a substrate for α -KG-dependent dioxygenases, such as PHDs, DNA hydroxylases and histone demethylases, whereas the TCA cycle intermediates fumarate and succinate inhibit PHDs via product-mediated inhibition. This stabilizes HIF-1 α and induces a transcriptional program that induces a shift from oxidative to glycolytic metabolism in T_H17 cells and M1 macrophages¹⁰¹. Yet another example is kynurenine, a tryptophan metabolite generated by indoleamine 2,3-dioxygenase (IDO), expressed by dendritic, T_{reg} and endothelial cells, which suppresses CTLs and T_H1 cells while stimulating T_H2 polarization and T_{reg} cell recruitment¹⁰². Finally, glutathione released by dendritic cells is cleaved to cysteine, which promotes T_{eff} cell proliferation; T_{reg} cells inhibit this signalling¹⁰³. Metabolic enzymes also engage in signalling in immune cells. For instance, GAPDH binds to the 3' untranslated region of cytokine mRNAs (such as IFN- γ) and thereby dampens their expression⁶⁰. Thus, when GAPDH is not supporting glycolysis for ATP production in T cells, it switches from its role as metabolic enzyme to an RNA-binding protein, thereby providing a metabolic checkpoint for T cells to couple effector status with glycolysis when sufficient nutrient is available⁶⁰.

by upregulating the expression of an anti-angiogenic VEGF-trap (that is, VEGFR-1) via unknown mechanisms²⁸. Fatty acid synthesis is relevant for angiogenesis, as its inhibition impairs capillary formation²⁹.

The hexosamine biosynthetic pathway is a side branch of glycolysis, which produces *N*-acetylglucosamine, an intermediate in *N*-linked and *O*-linked glycosylation. As this pathway uses glutamine, glucose, acetyl-CoA, uridine and ATP, it functions as a nutrient sensor. ECs use this pathway to glycosylate key angiogenic molecules, such as VEGFR-2, Notch and others. Interfering with glycosylation has contextual effects, but can nonetheless impair angiogenesis³⁰.

Dysregulated metabolism of ECs in diabetes

EC dysfunction

The endothelium modulates vascular tone, structure and homeostasis. Maladaptive metabolic changes in ECs contribute to vascular complications in diabetes. Functional disruption of the endothelium in type 2 diabetes is an

early event that promotes atherosclerosis and microangiopathy². Notably, ECs are highly glycolytic at normal plasma glucose levels (5.5 mM), but high glucose levels (25 mM) render glucose metabolism of ECs maladaptive. In an EC line, high glucose increases FAO while decreasing glucose oxidation, a contextual finding requiring confirmation in primary ECs³¹. Various metabolic changes have been proposed to promote vascular complications in diabetics, largely by increasing oxidative stress due to generation of various types of ROS (such as superoxide (O₂^{-•}) and hydrogen peroxide (H₂O₂)) and reactive nitrogen species (including nitric oxide (NO)) (Fig. 1b). First, high glucose levels cause mitochondrial dysfunction and increase the production of ROS via protein kinase C-dependent activation of NADPH-dependent oxidases, a major source of ROS in diabetic ECs³². Second, ROS detoxification is impaired in ECs, as high glucose decreases the production of NADPH for redox control by inhibiting glucose-6-phosphate-dehydrogenase (G6PD), a rate-limiting PPP enzyme. Third, ROS indirectly inactivate the glycolytic enzyme GAPDH, which lowers glycolysis³³. As a result of the reduced PPP and glycolytic flux, more glucose is shunted to the polyol pathway, which ECs use when exposed to high glucose in excess of what the glycolytic pathway can handle. Here, aldose reductase (AR) reduces glucose to sorbitol, which is further converted to fructose³⁴. As AR converts NADPH to NADP⁺, activation of this pathway depletes NADPH stores, which are necessary to maintain GSH levels for redox homeostasis, and thus promotes ROS accumulation. Sorbitol-derived fructose also generates the highly reactive 3-deoxyglucosone, which promotes the formation of advanced glycation end products (AGEs). Such glycated proteins and lipids contribute to vascular complications through formation of crosslinks between molecules in the basement membrane that alter the extracellular matrix. They also engage the receptor for AGEs (RAGE) on ECs, which induces vascular inflammation, leakage and further ROS production³⁵.

Fourth, ROS are generated when methylglyoxal is formed from the glycolytic metabolites dihydroxyacetone phosphate (DHAP) and glyceraldehyde-3-phosphate (GAP). The generation of methylglyoxal is enhanced in hyperglycaemic ECs because high glucose levels increase the formation of fructose (see above), which is metabolized to fructose-6-phosphate, and in turn to GAP and DHAP³⁶. Production of methylglyoxal is also increased via inactivation of GAPDH by ROS, which again increases GAP and DHAP levels (Fig. 1b). Elevated plasma levels of methylglyoxal in diabetic patients exacerbate EC dysfunction, and induce the formation of AGEs, oxidative stress and EC apoptosis, in addition to impairing vascular reactivity, altogether contributing to vascular complications and hypertension in diabetics³⁷.

Fifth, high glucose levels impair endothelial nitric oxide synthase (eNOS) activity through increased *O*-linked glycosylation. They also uncouple eNOS, which instead of generating NO, now produces O₂^{-•}. By inhibiting cytochrome *c* oxidase in the mitochondrial electron transport chain, these reactive species promote EC dysfunction^{38,39} (Fig. 1b). High glucose levels uncouple eNOS by increasing the levels of arginase, the enzyme that consumes L-arginine (referred to as arginine) necessary for the generation of NO (Fig. 1b).

Finally, ROS generated from high glucose exposure impairs angiogenesis via ligand-independent phosphorylation of VEGFR-2, leading to its reduced availability at the cell surface⁴⁰. This was verified in an *ob/ob* mouse model of diabetes, and the effect was reversed by treatment of diabetic mice with the antioxidant *N*-acetyl-L-cysteine⁴⁰. However, it should be cautioned that the precise contribution of oxidative stress to the development of vascular complications in diabetes remains debated. Also, despite all of this preclinical evidence, clinical antioxidant trials have not yielded overwhelmingly improved outcomes, highlighting the need for further study of these metabolic pathways⁴¹. Furthermore, studies on metabolic changes in diabetic ECs using metabolic tracing are warranted.

Ocular neovascularization

Diabetes impairs revascularization of ischaemic, inflamed and wounded tissues and impedes re-endothelialization of damaged vessels (such as after percutaneous coronary interventions), but also causes retinopathy characterized by overgrowth of leaky vessels inducing blindness². Vessel

overgrowth in diabetic retinopathy is stimulated by ischaemia, an angiogenic stimulus. The retinal ischaemia results from vessel hypoperfusion caused by multiple vascular anomalies, including vessel disintegration due to vascular cell death, leakage due to impaired junctional integrity and bleeding due to rupture of microaneurysms². Death of these vascular cells occurs as a result of excessive ROS production by an activated polyol pathway and mitochondrial dysfunction⁴². Increased glucose flux through the hexosamine biosynthetic pathway may also be involved, possibly via O-linked glycosylation of Akt⁴³. From a clinical perspective, it is worth noting that glycaemic control reduces micro- but not macrovascular complications, suggesting a distinctive role of EC metabolism in different vascular beds⁴⁴. Thus, it would be interesting to investigate whether the metabolism of micro- and macrovascular ECs differ.

Metabolism of healthy fibroblasts

Glycolysis and TCA cycle

Fibroblast metabolism has been studied only to a limited extent, but available evidence indicates that proliferating fibroblasts produce new biomass for replication, whereas quiescent fibroblasts generate biomass to replace oxidized lipids and degraded proteins, and to synthesize matrix proteins⁴⁵. Glycolysis is higher in proliferating fibroblasts than quiescent fibroblasts, but only by twofold, indicating that quiescent fibroblasts have high glycolytic needs for baseline homeostasis⁴⁵ (Fig. 2a, b). Quiescent and cycling fibroblasts incorporate glucose carbons in the TCA cycle at comparable rates⁴⁵. However, compared to proliferating fibroblasts in which glucose and glutamine metabolism in the TCA cycle is truncated at citrate, quiescent fibroblasts are able to use a fully operational TCA cycle. Quiescent and proliferating fibroblasts also differ in anaplerosis, a process that replenishes TCA cycle intermediates. Indeed, in quiescent fibroblasts, anaplerotic flux from pyruvate to oxaloacetate via pyruvate carboxylase ensures continuity of the TCA cycle, whereas proliferating fibroblasts primarily use glutamine for anaplerosis⁴⁵. Further, reductive carboxylation of α -KG by

mitochondrial isocitrate dehydrogenase (IDH)-2 is higher in quiescent fibroblasts, generating lipogenic acetyl-CoA for fatty acid synthesis⁴⁵ (Fig. 2a, b).

NADPH-producing pathways

Like cancer cells, proliferating fibroblasts rely on the PPP to generate ribose-5-phosphate and NADPH for biosynthesis, and incorporate glucose carbons via this pathway into nucleotides⁴⁵. By contrast, quiescent fibroblasts use the PPP to generate NADPH for redox homeostasis, with glucose carbons flowing back to glycolysis as GAP and fructose-6-phosphate⁴⁵. Quiescent fibroblasts also produce larger amounts of cytosolic NADPH by increasing the IDH-1 flux⁴⁵. The role of the PPP in redox balance is essential for quiescence, as its inhibition induces apoptosis of quiescent but not of cycling fibroblasts⁴⁵.

Metabolism of cancer-associated fibroblasts

Cancer-associated fibroblasts (CAFs) promote the growth and metastasis of cancers. A bidirectional crosstalk between cancer cells and CAFs has been identified, whereby cancer cells hijack CAFs to produce high-energy nutrients for their own use⁴⁶. In this crosstalk, production of ROS by cancer cells induces hypoxia-inducible factor 1 α (HIF-1 α) in neighbouring CAFs via inhibition of prolyl-hydroxylase domain (PHD) proteins. By inducing autophagy, HIF-1 α promotes the degradation of caveolin-1, a negative regulator of NO production. Excess production of NO then inhibits cytochrome c oxidase and leads to dysfunctional mitochondria, which further elevates ROS levels in CAFs⁴⁶. As CAFs remove dysfunctional mitochondria via mitophagy, they rely on enhanced glycolysis, leading to increased lactate production⁴⁶. CAFs also generate amino acids and ketone bodies, which together with lactate are used by cancer cells in more oxygenated areas as high-energy nutrients of oxidative metabolism to generate ATP^{47,48}. Via this metabolite crosstalk, CAFs thus feed nearby cancer cells. Proliferating CAFs also have increased rates of fatty acid synthesis, and blocking this process reduces their proliferation⁴⁹.

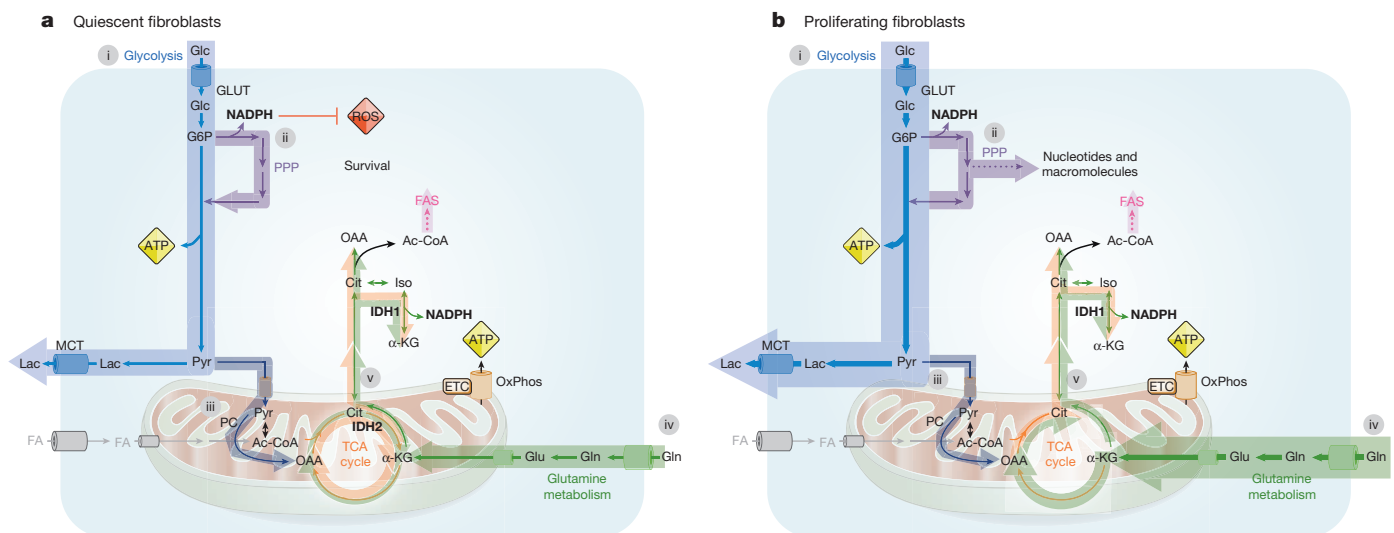


Figure 2 | Fibroblast metabolism. **a, b**, Scheme of the characteristic metabolic pathways in quiescent and proliferating fibroblasts. **a**, Quiescent fibroblasts: (i) have relatively high baseline glycolysis levels (blue). (ii) The PPP (purple) is important for redox homeostasis through enhanced production of NADPH (note that glucose carbons in the PPP flow back to glycolysis). (iii) Pyruvate carboxylase (dark blue) anaplerotically fills the TCA cycle with glucose carbons for glucose oxidation (orange), whereas (iv) glutamine is an anaplerotic carbon source for the TCA cycle (green; from glutamine to α -KG). (v) The TCA cycle proceeds in the forward (clockwise) direction. Quiescent fibroblasts have a high reductive carboxylation flux (green; from α -KG to citrate in anticlockwise reverse direction), generating acetyl-CoA for fatty acid synthesis (red) and NADPH for redox homeostasis by IDH1. Reductive carboxylation of α -KG, derived from glucose carbons, also occurs. **b**, Proliferating fibroblasts:

- (i) glycolysis (blue) is increased twofold in proliferating fibroblasts.
- (ii) The PPP is used to generate nucleotides and other macromolecules.
- (iii) Pyruvate carboxylase anaplerotically fills the TCA cycle with glucose carbons for glucose oxidation (dark blue), but (iv) glutamine is the major anaplerotic carbon source for the TCA cycle, consuming twice as much glutamine than quiescent cells (green; from glutamine to α -KG).
- (v) Proliferating fibroblasts have a truncated TCA cycle at citrate, preventing clockwise TCA cycling of glucose and glutamine carbons. However, compared to quiescent cells, they have a lower reductive carboxylation flux (green; from α -KG to citrate in the reverse direction anticlockwise), generating acetyl-CoA for FAS (red) and NADPH for redox homeostasis by IDH1. Iso, isocitrate; PC, pyruvate carboxylase; other abbreviations as in Fig. 1.

Immune cells

The innate immune system provides the first line of defence against infection by organisms bearing widely conserved pathogen motifs, whereas the adaptive immune system confers long-lasting antigen-specific protective immunity. The innate immune system is comprised of mast cells, macrophages, granulocytes (including basophils, eosinophils and neutrophils), dendritic cells and natural killer cells, whereas the adaptive immune system consists of B lymphocytes and T lymphocytes, including CD4⁺ T helper (T_H) cells and CD8⁺ cytotoxic T (CTL) cells. These immune cells contribute to tissue homeostasis in healthy conditions, but also influence tumour progression and metabolic disorders including type 2 diabetes^{50,51}. We will discuss T cells and macrophages, for which metabolism has been best characterized.

T-cell metabolism

The T-cell response to immunogens occurs in three steps: priming, expansion and contraction. Naive T (T_n) cells are quiescent, but after engagement by an antigen-presenting cell, T_n cells become activated, or primed. Upon activation of the T-cell receptor (TCR), T_n cells undergo an initial growth phase (increase in cell size) during the first 24 h, followed by clonal expansion and differentiation over the next 24–72 h. During activation, they differentiate into a heterogeneous population of subsets of T effector

(T_{eff}) cells. To divide every 6 to 8 h, and to produce effector molecules, T cells require increased amounts of energy and precursors for biomass production. Once the antigen is cleared, the T_{eff} cell population contracts and 90–95% of the activated T_{eff} cell pool dies, although a residual 5–10% of memory T cells (T_m in mice; T_{cm} (central memory) in humans) survives and provides immune memory; T_m cells can also differentiate from T_n cells⁵². By responding with greater speed and vigour to a repeat encounter with antigen, T_m cells protect the host without causing overt disease.

Metabolism during T-cell priming and expansion

T_n cells require minimal amounts of energy to prevent atrophy, survive and patrol the organism for antigen surveillance. Their ATP originates >90% from FAO and OxPhos, and to a lesser extent from glycolysis^{53,54} (Fig. 3b). T_n cells also have low levels of glutamine anaplerosis⁵⁵.

Activation of T cells induces a coordinated increase of glucose and amino acid transporters, and upregulates glucose and glutamine catabolism (Fig. 3b). Indeed, primed T lymphocytes increase glycolysis by 20–50-fold, even in normoxia⁵⁶. This increase stems partly from translocation of the glucose transporter GLUT1 to the plasma membrane and upregulation of glycolytic enzymes like hexokinase 2 and ADP-dependent glucokinase^{56–59}. Glucose consumption by the PPP is also increased, although it is unclear for which purpose^{54,58}. Aside from glycolysis, activated T cells also oxidize

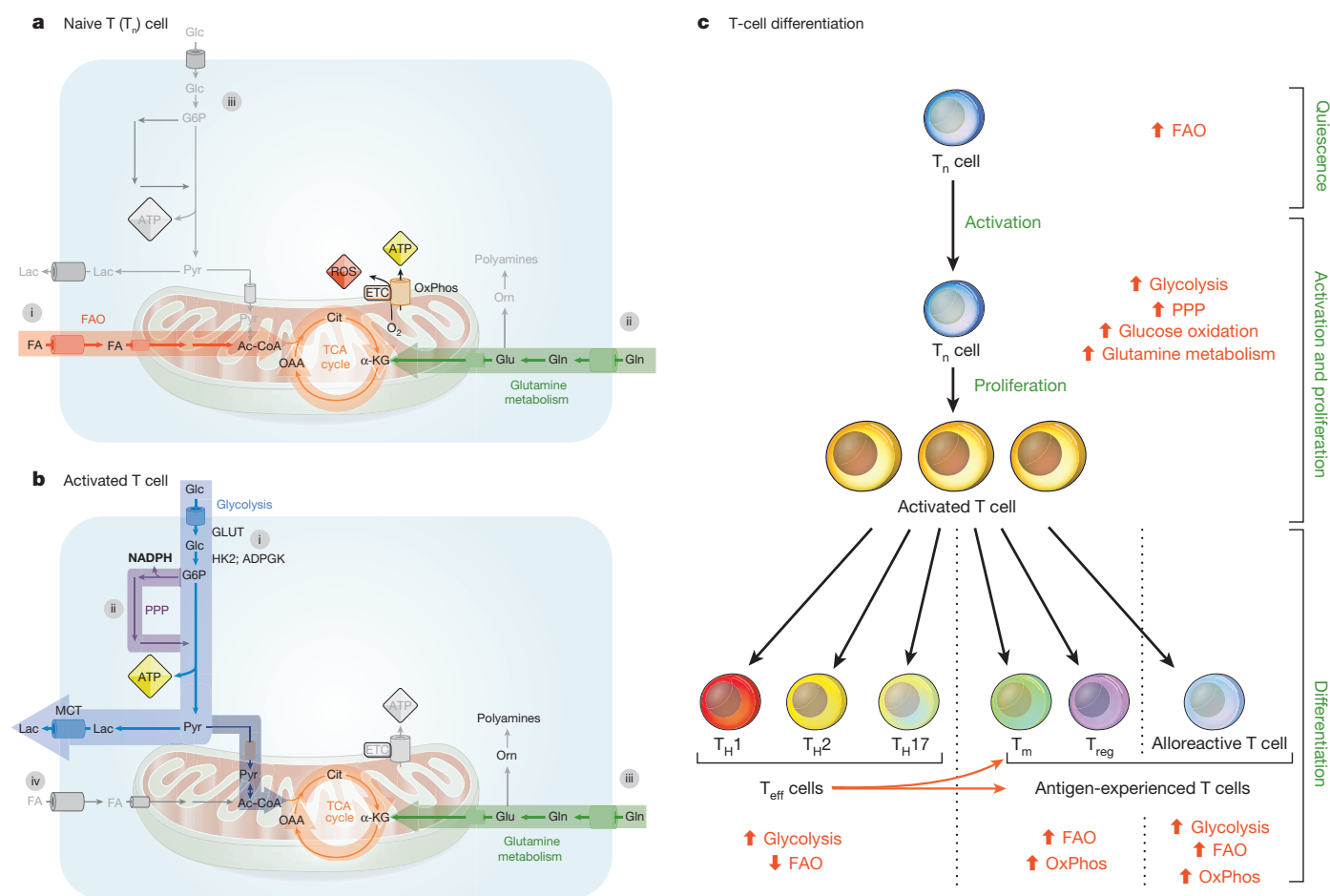


Figure 3 | T-cell metabolism. a, b, Metabolism of quiescent naive and activated T cells. **a,** (i) T_n cells rely primarily on FAO (red), (ii) baseline glutamine metabolism (green), (iii) but minimal glycolysis (grey). **b,** Upon activation, (i) they increase glucose uptake through GLUT1 and glycolysis by upregulation of hexokinase 2 (HK2) and ADP-dependent glucokinase (ADPGK) for ATP production. (ii) Increased PPP flux provides metabolites for redox homeostasis (NADPH) and proliferation (ribose-5-phosphate).

(iii) Increased uptake of glutamine fuels the TCA cycle via anaplerosis (green), (iv) while FAO levels are reduced. The TCA proceeds in clockwise direction (orange). **c,** Metabolic changes upon activation (priming) and differentiation of T cells towards their different subtypes. For reasons of clarity, differentiation of T_{reg} and T_m cells from T_n cells is not shown. Unlike T_m or T_n cells, T_{reg} cells continuously proliferate at moderate levels. Orn, ornithine; other abbreviations as in Fig. 1.

glutamine-derived α -KG⁶⁰. In addition, they metabolize glutamine via ornithine to mitogenic polyamines to promote their growth⁵⁸.

T-cell proliferation also relies on enhanced glucose and glutamine metabolism, and the importance of both pathways is exemplified by findings that deprivation of each nutrient blocks this process, even if other carbon sources are available^{56,58}. Although OxPhos can fuel T-cell proliferation⁶⁰, increased glucose uptake and glycolysis, not OxPhos, are necessary to activate CD4⁺ and CD8⁺ T cells to full effector status upon TCR activation, an effect relying in part on signalling by the glycolytic enzyme GAPDH in CD4⁺ T cells^{60,61} (Box 1). By inducing ROS signalling, OxPhos can also affect CD8⁺ T-cell expansion⁶².

Proliferating cells often elevate fatty acid synthesis (see above). Although oxidation of glutamine-derived α -KG in the TCA cycle is only a minor source of lipogenic acetyl-CoA in normal conditions (when it is primarily derived from glucose metabolism), α -KG can be converted to citrate via reductive carboxylation in conditions of hypoxic stress or mitochondrial dysfunction^{63,64}. T cells, which often operate in hypoxic wounds or inflamed tissues, preferentially use glutamine for biosynthesis of lipids by switching on reductive carboxylation⁶³. In agreement, TCR activation increases fatty acid synthesis and induces a redistribution of arachidonate from phosphatidylcholine to phosphatidylethanolamine species⁶⁵. Together, this coordinated metabolic switch promotes biosynthesis of proteins, nucleotides and phospholipids for rapid cell proliferation and effector status.

Metabolism during T-cell differentiation

Once T cells divide, they differentiate to different T-cell subsets, each switching on distinctive metabolic pathways. Activated CD4⁺ T cells differentiate to subsets of T_{eff} cells: T_{H1} cells mediate responses to intracellular pathogens, T_{H2} cells control responses to helminths, and T_{H17} cells are important in the defence against extracellular bacteria and fungi (Fig. 3c). T_{eff} cells upregulate glycolysis for anabolic purposes and suppress catabolic FAO for ATP production^{66,67} (Fig. 3c). T_{eff} cells also require glutamine for late effector status⁶⁸; however, the metabolic fate of glutamine oxidation remains unknown. Activated CD8⁺ CTLs also upregulate glucose and glutamine catabolism, but unlike CD4⁺ T cells, they do not increase OxPhos, making them more sensitive to glucose depletion^{53,54}.

Immune-suppressive regulatory T (T_{reg}) cells, another differentiated cell type, rely on FAO for ATP production^{67,69}. Interestingly, these cells do not exhibit a sudden proliferation burst like T_{eff} cells, but continuously proliferate at moderate levels. Whether this might explain their dependence on FAO rather than glycolysis is unknown. T_{reg} cells also differ from T_{H17} cells in their need for amino acid metabolism. Indeed, depletion of extracellular amino acids suppresses T_{H17} cell differentiation, while favouring T_{reg} cell development and T-cell anergy (a long-term state of hyporesponsiveness in response to suboptimal stimulation)⁷⁰.

T_m cells adopt a metabolic profile comparable to that of T_n cells, but differ by increased mitochondrial mass and reserve respiratory capacity, which prepares them for rapid mitochondrial ATP production upon TCR re-engagement, and provides a bioenergetic advantage upon secondary antigen exposure⁷¹. Hence, specific metabolic programs in different immune cells are critical for multiple aspects and during specific phases of the immune response, with engagement of glycolysis and FAO during acute and more chronic antigen exposure, respectively.

The metabolic phenotype of activated T cells is influenced by the microenvironment. Indeed, during graft versus host disease (GVHD), allogenic T_{eff} cells increase not only glycolysis but also fatty acid uptake and FAO to support activation and proliferation *in vivo*⁷². Presumably, the prolonged presentation of antigens to T cells combined with other factors (for example, the type of antigen-presenting cell) during GVHD drives these cells to generate ATP as efficiently as possible by using oxidative metabolism as well (Fig. 3c). This is different from a condition whereby only limited amounts of antigens are presented for a finite period with minimal amounts of systemic inflammation, which T cells can handle by upregulating glycolysis alone. Signalling by metabolites and metabolic enzymes, and metabolite crosstalk, also regulate T-cell growth, differentiation and function (Box 2).

BOX 2

Metabolite crosstalks

Metabolite exchange between stromal and parenchymal cells is common and essential for the function of numerous tissues. For instance, extracellular lactate produced by ECs functions as a dynamic vasoactive signal for pericytes. Indeed, when cellular energy availability is low, lactate causes vasodilation to promote fuel supply, but when energy supplies are abundant, lactate acts as a vasoconstrictor¹⁰⁴. This glycolytic metabolite, produced by cancer cells, also promotes angiogenesis, inhibits cytotoxic T-cell responses, and causes chronic inflammation through enhanced IL-17A secretion, which activates T_{H17} cells¹⁰². As T_{eff} cells consume glucose intensely during early inflammation, glucose deprivation might promote T_{reg} cell differentiation to dampen inflammation⁶⁷. Further, acidification of the tumour microenvironment by lactate increases arginase 1 levels in macrophages, which limits the anti-tumour immune response (see text)¹⁰⁵. This effect can be abrogated by the glucose oxidation inhibitor dichloroacetate¹⁰⁵. An analogous metabolic cooperation exists between cancer-activated adipocytes and cancer cells, where cancer cells reprogram the metabolism of adipocytes to become catabolic, so that they release fatty acids as fuel for cancer cells¹⁰⁶.

Macrophage metabolism

M1-like macrophages

Macrophages are terminally differentiated cells that do not need energy and biomass to proliferate but instead to sustain a high phagocytic and secretory activity. Macrophages are polarized to distinct phenotypes⁷³. When activated by interferon (IFN)- γ alone or in concert with microbial stimuli (like LPS) or pro-inflammatory cytokines (like tumour necrosis factor (TNF)- α), they acquire an inflammatory phenotype, and are termed 'classically activated' M1 macrophages. These cells have a high microbicidal activity, as well as enhanced production of pro-inflammatory cytokines and ROS. M1 macrophages increase glycolysis by inducing the expression of the pro-glycolytic PFKFB3 isoform⁷⁴ (Fig. 4a). Their dependence on glycolysis rather than mitochondrial ATP production, even in normoxia, offers M1 macrophages energetic advantages in hypoxic regions⁷⁵.

LPS has various effects on macrophage metabolism: first, it increases levels of the TCA intermediate succinate, which upregulates IL-1 β expression via activation of HIF-1 α (succinate inhibits PHDs that negatively regulate HIF-1 α)⁷⁶. Second, it increases succinylation of metabolic proteins, but the precise function of this modification remains unknown⁷⁶. Third, LPS increases the expression of immunoresponsive gene 1 (IRG1), an enzyme producing itaconic acid⁷⁷ (Fig. 4a). Because itaconic acid inhibits isocitrate lyase, the key enzyme of the glyoxylate shunt in pathogens, this metabolite exerts antimicrobial activity during bacterial infections⁷⁷.

A key feature of M1 macrophages is their ability to produce high levels of mitochondrial ROS to kill phagocytosed bacteria. IRG1 enhances FAO and OxPhos to generate bactericidal mitochondrial ROS inside phagosomes containing ingested bacteria⁷⁸ (Fig. 4a). M1 macrophages also release ROS and reactive nitrogen species (including NO) in phagosomes for killing pathogens. ROS and NO for this purpose are generated by NADPH oxidase and inducible nitric oxide synthase (iNOS), respectively (Fig. 4a). Both enzymes require NADPH, which is produced in the PPP as well as by malic enzyme. The latter NADPH-producing mechanism is induced in response to LPS. In brief, LPS increases transport of mitochondrial citrate to the cytosol, where it is converted to oxaloacetate and then to pyruvate by malic enzyme, thereby generating NADPH⁷⁹. To protect themselves against ROS leaking into the cytosol during such high oxidative burst, M1 macrophages also use NADPH from the PPP for the production of the antioxidant GSH (Fig. 4b).

To phagocytose, a large turnover of lipids and generation of membrane-rich filopodia and organelles are required. Macrophages therefore increase phospholipid synthesis and switch from cholesterol to phosphatidylcholine

remarkable, as Notch is the most potent pro-stalk signal known to date. Thus, PFKFB3-driven glycolysis regulates the plasticity of EC phenotypes, and not only genetic but also metabolic signals regulate vessel sprouting.

Immune cells

Metabolism also controls T-cell lineage choices. In accordance with the importance of glycolysis for T_{eff} cells, activation of glycolysis by transgenic expression of GLUT1 or glycolytic genes augments T-cell activation, results in the accumulation of readily activated T_m -like cells with signs of autoimmunity in aged mice⁵⁶, and drives $CD8^+$ T cells towards a terminally differentiated, more short-lived state⁸⁸. Conversely, blocking glycolysis via pharmacological or transgenic means impairs T_H17 cell proliferation and differentiation while promoting T_{reg} cells, thereby protecting mice from autoimmune inflammation^{58,66}. Moreover, inhibiting glucose metabolism preserves the formation of long-lived $CD8^+$ T_m cells *in vivo*⁸⁸, while rendering T cells anergic⁵⁴.

Amino acid and fatty acid metabolism are also important, as inhibition of glutaminase impairs T_{eff} -cell proliferation⁵⁸, whereas FAO blockade inhibits T_{reg} differentiation⁶⁷ and the survival of alloreactive T cells⁷². Conversely, the short-chain fatty acid butyrate, produced by commensal microorganisms, promotes T_{reg} -cell development in the colon, thereby protecting mice against colitis^{89–91}. Also, by mediating leucine uptake, the amino acid transporter SLC7A5 induces $CD4^+$ and $CD8^+$ T-cell differentiation via activation of mammalian target of rapamycin complex 1 (mTORC1), glycolysis and glutamine metabolism, thereby functioning as a metabolic checkpoint⁹². Macrophage differentiation is also determined by metabolism. Indeed, pretreatment of macrophage precursors with respiration or FAO blockers abrogates M2 activation, without affecting the M1 phenotype⁸¹. Aside from a role in metabolism, metabolites and metabolic enzymes modulate the plasticity of immune cells by affecting gene transcription, signalling pathways or epigenetic programs (Box 2).

Metabolic commonalities and differences

This overview highlights that ECs, fibroblasts and immune cells both share and differ in their use of metabolic pathways, although metabolism is incompletely characterized in some of these cell types. For instance, proliferating endothelial stalk cells, fibroblasts, T_H1 cells, T_H2 cells and T_H17 cells all rely on increased glycolysis, glutamine metabolism and fatty acid synthesis to generate the needed amount of ATP and new biomass for cell division. Endothelial tip cells, T_{eff} cells and activated M1 macrophages also rely on glycolysis, but use it instead to achieve full effector status. By contrast, T_m cells, T_{reg} cells, alloreactive T cells and M2 macrophages rely primarily on FAO to generate ATP more efficiently. Oxidative metabolism seems to be best suited to meet the bioenergetic demands for the long-term performance of these cells. Reductive carboxylation of α -KG for fatty acid synthesis has only been identified to date in M1 macrophages and fibroblasts, but is probably also relevant for other non-transformed cell types.

The PPP is of vital importance for many cells, but is used for distinctive purposes in different stromal and immune cell types. For instance, the PPP is essential for biosynthesis of macromolecules in proliferating fibroblasts and probably also in other dividing cells (ECs, T cells), although evidence is lacking so far. The PPP and other NADPH-generating pathways (including FAO) are used by M1 macrophages for the production of ROS to kill pathogens. To buffer this oxidative burst, they must also use the PPP for redox homeostasis. Quiescent fibroblasts are unusually sensitive to redox balance offered by the PPP. ROS have a signalling role in ECs and T cells, unlike their lethal bactericidal effects in M1 macrophages.

Quiescent ECs and fibroblasts differ from T_n cells in their needs for glycolysis during homeostatic maintenance. This may relate to the fact that quiescent T_n cells are largely inactive, awaiting activation by foreign antigen, whereas quiescent ECs and fibroblasts must continuously maintain the vascular barrier, transport molecules, synthesize matrix, replenish degraded macromolecules and perform homeostatic activities. Unlike other stromal and immune cells, macrophages do not use metabolism for division, but to resynthesize degraded biomass and achieve full effector

status. Overall, stromal and immune cells have adopted their metabolism to optimally fulfil their distinctive role as support cells in particular conditions, in which each of these stromal and immune cells must fulfil specialized functions.

Outstanding questions and perspectives

We suspect that the studies discussed here merely scratch the surface of the metabolic control of stromal and immune cells. Multiple questions require an answer in the future. A priority for EC and fibroblast metabolism research is to establish a road map of additional metabolic pathways that have not been studied to date. It will be equally fruitful to study how stromal and immune cell types metabolically communicate with each other and other (cancer) cells, and whether pathways or signalling properties of metabolic enzymes, characterized in cancer cells, also exist in stromal and immune cells. Analysing the metabolism of these cell types *in vivo*, in preclinical animal models and ultimately in patients is an important yet challenging translational priority for the future. Furthermore, little is known about the metabolism of various stromal cell types (for example, arterial, venous, lymphatic or tumour ECs; pericytes and smooth muscle cells; bone marrow mesenchymal cells, hepatic stellate cells) and other immune cells (for example, neutrophils, basophils, monocytes, dendritic cells) or of their progenitors, and whether metabolism determines their phenotypic identity. For instance, do tip and stalk ECs have distinct metabolic signatures (besides a difference in glycolysis), and do pro-tip or pro-stalk cell signals determine specific metabolic pathways?

Another question is whether stromal and immune cell metabolism differs from cancer cell metabolism. It has been postulated that the unique properties of cancer cell metabolism would offer opportunities to selectively target cancer cells. Although this seems plausible for tumours in which metabolism is changed by genetic alteration or overexpression of metabolic targets, it is unknown whether blocking metabolic changes more commonly seen in other tumours will target cancer cells selectively. Although metabolic similarities in cancer and stromal/immune cells may enhance anti-cancer therapies targeting a common metabolic pathway, possible consequences on healthy stromal cells, which also rely on these pathways, will need to be evaluated.

Finally, what are the translational consequences of these insights into stromal and immune cell metabolism for malignant, metabolic or inflammatory disease? Will it be possible to target particular stromal or immune cells, and target only activated or proliferating cells without affecting quiescent cells? Reports that blocking glycolysis protects mice from autoimmune inflammation by impairing T_H17 cell proliferation and differentiation, while favouring T_{reg} cell differentiation, suggest that this might be possible^{58,66}. Another example is the transient and partial reduction of glycolysis by PFKFB3 blockade, which is sufficient to reduce pathological angiogenesis, without affecting pre-established vessels or evoking overt toxicity or metabolic compensation¹⁴.

In preclinical models, lowering glycolysis improves psoriasis⁹³, protects against autoimmune inflammation^{58,66}, reduces ocular neovascularization (a common complication in diabetics)¹⁴ and decreases tumour growth by inhibiting the growth of cancer cells and probably of ECs, fibroblasts and T cells⁹⁴. In addition, glycolytic blockade enhances the anti-angiogenic effect of VEGF inhibitors, indicating that targeting metabolism can amplify available treatments¹⁴. The preclinical finding that inhibition of AR improves vascular complications in diabetes⁹⁵ has been translated to the clinical situation, and treatment of human diabetics with the AR inhibitor epalrestat prevents further progression of retinopathy⁹⁶. These examples illustrate the therapeutic potential of targeting metabolism in stromal and immune cells, and should provide an incentive to further characterize metabolism in these cells.

Received 2 December 2013; accepted 8 April 2014.

1. Carmeliet, P. & Jain, R. K. Molecular mechanisms and clinical applications of angiogenesis. *Nature* **473**, 298–307 (2011).
2. Rask-Madsen, C. & King, G. L. Vascular complications of diabetes: mechanisms of injury and protective factors. *Cell Metab.* **17**, 20–33 (2013).

3. Junttila, M. R. & de Sauvage, F. J. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature* **501**, 346–354 (2013).
4. Galluzzi, L., Kepp, O., Vander Heiden, M. G. & Kroemer, G. Metabolic targets for cancer therapy. *Nature Rev. Drug Discov.* **12**, 829–846 (2013).
Provides a comprehensive overview of metabolic targets that may also be suitable for use in stromal and immune cells.
5. Schulze, A. & Harris, A. L. How cancer metabolism is tuned for proliferation and vulnerable to disruption. *Nature* **491**, 364–373 (2012).
6. Lunt, S. Y. & Vander Heiden, M. G. Aerobic glycolysis: meeting the metabolic requirements of cell proliferation. *Annu. Rev. Cell Dev. Biol.* **27**, 441–464 (2011).
7. DeBerardinis, R. J. & Thompson, C. B. Cellular metabolism and disease: what do metabolic outliers teach us? *Cell* **148**, 1132–1144 (2012).
Refs 5–7 highlight emerging concepts of the role of metabolism in cell proliferation and cancer.
8. Casazza, A. *et al.* Tumor stroma: a complexity dictated by the hypoxic tumor microenvironment. *Oncogene* **33**, 1743–1754 (2014).
9. Potente, M., Gerhardt, H. & Carmeliet, P. Basic and therapeutic aspects of angiogenesis. *Cell* **146**, 873–887 (2011).
10. De Bock, K. *et al.* Role of PFKFB3-driven glycolysis in vessel sprouting. *Cell* **154**, 651–663 (2013).
The first description of how metabolism can alter tip cell competitiveness in endothelial sprouting in vitro and in vivo.
11. Barth, E., Stamm, B., Speiser, B. & Schaper, J. Ultrastructural quantitation of mitochondria and myofibrils in cardiac muscle from 10 different animal species including man. *J. Mol. Cell. Cardiol.* **24**, 669–681 (1992).
12. Quintero, M., Colombo, S. L., Godfrey, A. & Moncada, S. Mitochondria as signaling organelles in the vascular endothelium. *Proc. Natl Acad. Sci. USA* **103**, 5379–5384 (2006).
13. Chung, S. J. *et al.* Pyruvate protection against endothelial cytotoxicity induced by blockade of glucose uptake. *J. Biochem. Mol. Biol.* **37**, 239–245 (2004).
14. Schoors, S. *et al.* Partial and transient reduction of glycolysis by PFKFB3 blockade reduces pathological angiogenesis. *Cell Metab.* **19**, 37–48 (2014).
Demonstrates that targeting endothelial cell metabolism can inhibit pathological angiogenesis.
15. Locasale, J. W. *et al.* Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. *Nature Genet.* **43**, 869–874 (2011).
16. Gatenby, R. A. & Gillies, R. J. Why do cancers have high aerobic glycolysis? *Nature Rev. Cancer* **4**, 891–899 (2004).
17. Mertens, S., Noll, T., Spahr, R., Krutzfeldt, A. & Piper, H. M. Energetic response of coronary endothelial cells to hypoxia. *Am. J. Physiol.* **258**, H689–H694 (1990).
18. Leopold, J. A. *et al.* Aldosterone impairs vascular reactivity by decreasing glucose-6-phosphate dehydrogenase activity. *Nature Med.* **13**, 189–197 (2007).
19. Leopold, J. A., Zhang, Y. Y., Scribner, A. W., Stanton, R. C. & Loscalzo, J. Glucose-6-phosphate dehydrogenase overexpression decreases endothelial cell oxidant stress and increases bioavailable nitric oxide. *Arterioscler. Thromb. Vasc. Biol.* **23**, 411–417 (2003).
20. Spolarics, Z., Lang, C. H., Bagby, G. J. & Spitzer, J. J. Glutamine and fatty acid oxidation are the main sources of energy for Kupffer and endothelial cells. *Am. J. Physiol.* **261**, G185–G190 (1991).
21. Polet, F. & Feron, O. Endothelial cell metabolism and tumour angiogenesis: glucose and glutamine as essential fuels and lactate as the driving force. *J. Intern. Med.* **273**, 156–165 (2013).
22. Unterluggauer, H. *et al.* Premature senescence of human endothelial cells induced by inhibition of glutamine. *Biogerontology* **9**, 247–259 (2008).
23. Wu, G., Haynes, T. E., Li, H. & Meininger, C. J. Glutamine metabolism in endothelial cells: ornithine synthesis from glutamine via pyrroline-5-carboxylate synthase. *Comp. Biochem. Physiol. Physiol.* **126**, 115–123 (2000).
24. Carracedo, A., Cantley, L. C. & Pandolfi, P. P. Cancer metabolism: fatty acid oxidation in the limelight. *Nature Rev. Cancer* **13**, 227–232 (2013).
A comprehensive overview of FAO-related pathways in cancer that may provide putative targets for stromal and immune cell metabolism.
25. Dagher, Z., Ruderman, N., Tornheim, K. & Ido, Y. Acute regulation of fatty acid oxidation and AMP-activated protein kinase in human umbilical vein endothelial cells. *Circ. Res.* **88**, 1276–1282 (2001).
26. Schug, Z. T., Frezza, C., Galbraith, L. C. & Gottlieb, E. The music of lipids: how lipid composition orchestrates cellular behaviour. *Acta Oncol.* **51**, 301–310 (2012).
Illustrates the importance of lipid composition on mitochondrial structure, dynamics and bioenergetics in relation to cellular function.
27. Fang, L. *et al.* Control of angiogenesis by AIBP-mediated cholesterol efflux. *Nature* **498**, 118–122 (2013).
28. Avraham-David, I. *et al.* ApoB-containing lipoproteins regulate angiogenesis by modulating expression of VEGF receptor 1. *Nature Med.* **18**, 967–973 (2012).
29. Seguin, F. *et al.* The fatty acid synthase inhibitor orlistat reduces experimental metastases and angiogenesis in B16-F10 melanomas. *Br. J. Cancer* **107**, 977–987 (2012).
30. Merchan, J. R. *et al.* Antiangiogenic activity of 2-deoxy-D-glucose. *PLoS ONE* **5**, e13699 (2010).
31. Koziel, A., Woyda-Ploszczyca, A., Kicinska, A. & Jarmuszkiewicz, W. The influence of high glucose on the aerobic metabolism of endothelial EA.hy926 cells. *Pflugers Arch.* **464**, 657–669 (2012).
32. Brandes, R. P., Weissmann, N. & Schroder, K. Redox-mediated signal transduction by cardiovascular Nox NADPH oxidases. *J. Mol. Cell. Cardiol.* **73**, 70–79 (2014).
33. Du, X. *et al.* Inhibition of GAPDH activity by poly(ADP-ribose) polymerase activates three major pathways of hyperglycemic damage in endothelial cells. *J. Clin. Invest.* **112**, 1049–1057 (2003).
34. Tang, W. H., Martin, K. A. & Hwa, J. Aldose reductase, oxidative stress, and diabetic mellitus. *Front Pharmacol.* **3**, 87 (2012).
35. Manigrasso, M. B., Juranek, J., Ramasamy, R. & Schmidt, A. M. Unlocking the biology of RAGE in diabetic microvascular complications. *Trends Endocrinol. Metab.* **25**, 15–22 (2014).
36. Liu, J. *et al.* Aldolase B knockdown prevents high glucose-induced methylglyoxal overproduction and cellular dysfunction in endothelial cells. *PLoS ONE* **7**, e41495 (2012).
37. Sena, C. M. *et al.* Methylglyoxal promotes oxidative stress and endothelial dysfunction. *Pharmacol. Res.* **65**, 497–506 (2012).
38. Forstermann, U. & Sessa, W. C. Nitric oxide synthases: regulation and function. *Eur. Heart J.* **33**, 829–837 (2012).
39. Beleznaï, T. & Bagi, Z. Activation of hexosamine pathway impairs nitric oxide (NO)-dependent arteriolar dilations by increased protein O-GlcNAcylation. *Vascul. Pharmacol.* **56**, 115–121 (2012).
40. Warren, C. M., Ziyad, S., Briot, A., Der, A. & Iruela-Arispe, M. L. A ligand-independent VEGFR2 signaling pathway limits angiogenic responses in diabetes. *Sci. Signal.* **7**, ra1 (2014).
Mechanistic evidence of the anti-angiogenic effect of high glucose and ROS in diabetes.
41. Hill, M. F. Emerging role for antioxidant therapy in protection against diabetic cardiac complications: experimental and clinical evidence for utilization of classic and new antioxidants. *Curr. Cardiol. Rev.* **4**, 259–268 (2008).
42. Shah, G. N. *et al.* Pharmacological inhibition of mitochondrial carbonic anhydrases protects mouse cerebral pericytes from high glucose-induced oxidative stress and apoptosis. *J. Pharmacol. Exp. Ther.* **344**, 637–645 (2013).
43. Luo, B., Soesanto, Y. & McClain, D. A. Protein modification by O-linked GlcNAc reduces angiogenesis by inhibiting Akt activity in endothelial cells. *Arterioscler. Thromb. Vasc. Biol.* **28**, 651–657 (2008).
44. Dluhy, R. G. & McMahon, G. T. Intensive glycemic control in the ACCORD and ADVANCE trials. *N. Engl. J. Med.* **358**, 2630–2633 (2008).
45. Lemons, J. M. *et al.* Quiescent fibroblasts exhibit high metabolic activity. *PLoS Biol.* **8**, e1000514 (2010).
Provides the first comprehensive description of the major metabolic differences between quiescent and proliferating fibroblasts.
46. Pavlides, S. *et al.* Warburg meets autophagy: cancer-associated fibroblasts accelerate tumor growth and metastasis via oxidative stress, mitophagy, and aerobic glycolysis. *Antioxid. Redox Signal.* **16**, 1264–1284 (2012).
47. Lisanti, M. P. *et al.* Understanding the “lethal” drivers of tumor-stroma co-evolution: emerging role(s) for hypoxia, oxidative stress and autophagy/mitophagy in the tumor micro-environment. *Cancer Biol. Ther.* **10**, 537–542 (2010).
48. Chaudhri, V. K. *et al.* Metabolic alterations in lung cancer-associated fibroblasts correlated with increased glycolytic metabolism of the tumor. *Mol. Cancer Res.* **11**, 579–592 (2013).
49. Santolla, M. F. *et al.* G protein-coupled estrogen receptor mediates the up-regulation of fatty acid synthase induced by 17 β -estradiol in cancer cells and cancer-associated fibroblasts. *J. Biol. Chem.* **287**, 43234–43245 (2012).
50. Gajewski, T. F., Schreiber, H. & Fu, Y. X. Innate and adaptive immune cells in the tumor microenvironment. *Nature Immunol.* **14**, 1014–1022 (2013).
51. Eheim, A., Medrikova, D. & Herzig, S. Immune cells and metabolic dysfunction. *Semin. Immunopathol.* **36**, 13–25 (2014).
52. Jameson, S. C. & Masopust, D. Diversity in T cell memory: an embarrassment of riches. *Immunity* **31**, 859–871 (2009).
53. Wahl, D. R., Byersdorfer, C. A., Ferrara, J. L., Opipari, A. W., Jr & Glick, G. D. Distinct metabolic programs in activated T cells: opportunities for selective immunomodulation. *Immunol. Rev.* **249**, 104–115 (2012).
54. MacIver, N. J., Michalek, R. D. & Rathmell, J. C. Metabolic regulation of T lymphocytes. *Annu. Rev. Immunol.* **31**, 259–283 (2013).
55. Pearce, E. L., Poffenberger, M. C., Chang, C. H. & Jones, R. G. Fueling immunity: insights into metabolism and lymphocyte function. *Science* **342**, 1242454 (2013).
Refs 50–55 highlight emerging concepts of metabolism and its importance in T-cell function and development.
56. Jacobs, S. R. *et al.* Glucose uptake is limiting in T cell activation and requires CD28-mediated Akt-dependent and independent pathways. *J. Immunol.* **180**, 4476–4486 (2008).
57. Kamiński, M. M. *et al.* T cell activation is driven by an ADP-dependent glucokinase linking enhanced glycolysis with mitochondrial reactive oxygen species generation. *Cell Rep* **2**, 1300–1315 (2012).
58. Wang, R. *et al.* The transcription factor Myc controls metabolic reprogramming upon T lymphocyte activation. *Immunity* **35**, 871–882 (2011).
Study describing the Myc-dependent global metabolic transcriptome driving metabolic reprogramming in activated T cells.
59. MacIver, N. J. *et al.* The liver kinase B1 is a central regulator of T cell development, activation, and metabolism. *J. Immunol.* **187**, 4187–4198 (2011).
60. Chang, C. H. *et al.* Posttranscriptional control of T cell effector function by aerobic glycolysis. *Cell* **153**, 1239–1251 (2013).
Demonstrates translational control of T-cell activation through GAPDH binding to IFN- γ mRNA.
61. Cham, C. M., Driessens, G., O’Keefe, J. P. & Gajewski, T. F. Glucose deprivation inhibits multiple key gene expression events and effector functions in CD8 $^{+}$ T cells. *Eur. J. Immunol.* **38**, 2438–2450 (2008).
62. Sena, L. A. *et al.* Mitochondria are required for antigen-specific T cell activation through reactive oxygen species signaling. *Immunity* **38**, 225–236 (2013).
63. Metallo, C. M. *et al.* Reductive glutamine metabolism by IDH1 mediates lipogenesis under hypoxia. *Nature* **481**, 380–384 (2012).

Shows the importance of reductive carboxylation of α -KG to produce acetyl-CoA for *de novo* lipid synthesis.

64. Wise, D. R. *et al.* Hypoxia promotes isocitrate dehydrogenase-dependent carboxylation of α -ketoglutarate to citrate to support cell growth and viability. *Proc. Natl Acad. Sci. USA* **108**, 19611–19616 (2011).
65. Robichaud, P. P., Boulay, K., Munganyiki, J. E. & Surette, M. E. Fatty acid remodeling in cellular glycerophospholipids following the activation of human T cells. *J. Lipid Res.* **54**, 2665–2677 (2013).
66. Shi, L. Z. *et al.* HIF1 α -dependent glycolytic pathway orchestrates a metabolic checkpoint for the differentiation of T_H17 and T_{reg} cells. *J. Exp. Med.* **208**, 1367–1376 (2011).
67. Michalek, R. D. *et al.* Cutting edge: distinct glycolytic and lipid oxidative metabolic programs are essential for effector and regulatory CD4⁺ T cell subsets. *J. Immunol.* **186**, 3299–3303 (2011).
68. Carr, E. L. *et al.* Glutamine uptake and metabolism are coordinately regulated by ERK/MAPK during T lymphocyte activation. *J. Immunol.* **185**, 1037–1044 (2010).
69. Wang, R. & Green, D. R. Metabolic checkpoints in activated T cells. *Nature Immunol.* **13**, 907–915 (2012).
- Summarizes metabolic requirements for T-cell activation.**
70. Sundrud, M. S. *et al.* Halofuginone inhibits T_H17 cell differentiation by activating the amino acid starvation response. *Science* **324**, 1334–1338 (2009).
71. van der Windt, G. J. *et al.* Mitochondrial respiratory capacity is a critical regulator of CD8⁺ T cell memory development. *Immunity* **36**, 68–78 (2012).
72. Byersdorfer, C. A. *et al.* Effector T cells require fatty acid metabolism during murine graft-versus-host disease. *Blood* **122**, 3230–3237 (2013).
- Shows the ability of T_{eff} to adopt a different metabolic phenotype inflicted by external cues.**
73. Mantovani, A., Biswas, S. K., Galdiero, M. R., Sica, A. & Locati, M. Macrophage plasticity and polarization in tissue repair and remodelling. *J. Pathol.* **229**, 176–185 (2013).
74. Rodríguez-Prados, J. C. *et al.* Substrate fate in activated macrophages: a comparison between innate, classic, and alternative activation. *J. Immunol.* **185**, 605–614 (2010).
75. Nizet, V. & Johnson, R. S. Interdependence of hypoxic and innate immune responses. *Nature Rev. Immunol.* **9**, 609–617 (2009).
76. Tannahill, G. M. *et al.* Succinate is an inflammatory signal that induces IL-1 β through HIF-1 α . *Nature* **496**, 238–242 (2013).
- Identification of a crucial role for succinate in innate immune signalling, through enhancement of IL-1 β production during inflammation.**
77. Michelucci, A. *et al.* Immune-responsive gene 1 protein links metabolism to immunity by catalyzing itaconic acid production. *Proc. Natl Acad. Sci. USA* **110**, 7820–7825 (2013).
78. Hall, C. J. *et al.* *Immunoresponsive gene 1* augments bactericidal activity of macrophage-lineage cells by regulating β -oxidation-dependent mitochondrial ROS production. *Cell Metab.* **18**, 265–278 (2013).
79. O'Neill, L. A. A critical role for citrate metabolism in LPS signalling. *Biochem. J.* **438**, e5–e6 (2011).
80. Ecker, J. *et al.* Induction of fatty acid synthesis is a key requirement for phagocytic differentiation of human monocytes. *Proc. Natl Acad. Sci. USA* **107**, 7817–7822 (2010).
81. Vats, D. *et al.* Oxidative metabolism and PGC-1 β attenuate macrophage-mediated inflammation. *Cell Metab.* **4**, 13–24 (2006).
82. O'Neill, L. A. & Hardie, D. G. Metabolism of inflammation limited by AMPK and pseudo-starvation. *Nature* **493**, 346–355 (2013).
83. Sene, A. *et al.* Impaired cholesterol efflux in senescent macrophages promotes age-related macular degeneration. *Cell Metab.* **17**, 549–561 (2013).
84. Haschemi, A. *et al.* The sedoheptulose kinase CARKL directs macrophage polarization through control of glucose metabolism. *Cell Metab.* **15**, 813–826 (2012).
85. Biswas, S. K. & Mantovani, A. Orchestration of metabolism by macrophages. *Cell Metab.* **15**, 432–437 (2012).
- An overview of the metabolic adaptations underlying macrophage plasticity and polarization and the critical use of metabolism in their function.**
86. Mills, C. D. M1 and M2 macrophages: oracles of health and disease. *Crit. Rev. Immunol.* **32**, 463–488 (2012).
87. Popovic, P. J., Zeh, H. J. III & Ochoa, J. B. Arginine and immunity. *J. Nutr.* **137**, 1681S–1686S (2007).
88. Sukumar, M. *et al.* Inhibiting glycolytic metabolism enhances CD8⁺ T cell memory and antitumor function. *J. Clin. Invest.* **123**, 4479–4488 (2013).
89. Arpaia, N. *et al.* Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature* **504**, 451–455 (2013).
90. Furusawa, Y. *et al.* Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* **504**, 446–450 (2013).
91. Smith, P. M. *et al.* The microbial metabolites, short-chain fatty acids, regulate colonic T_{reg} cell homeostasis. *Science* **341**, 569–573 (2013).
92. Sinclair, L. V. *et al.* Control of amino-acid transport by antigen receptors coordinates the metabolic reprogramming essential for T cell differentiation. *Nature Immunol.* **14**, 500–508 (2013).
93. Telang, S. *et al.* Small molecule inhibition of 6-phosphofructo-2-kinase suppresses T cell activation. *J. Transl. Med.* **10**, 95 (2012).
94. Clem, B. *et al.* Small-molecule inhibition of 6-phosphofructo-2-kinase activity suppresses glycolytic flux and tumor growth. *Mol. Cancer Ther.* **7**, 110–120 (2008).
95. Obrosova, I. G. & Kador, P. F. Aldose reductase/polyol inhibitors for diabetic retinopathy. *Curr. Pharm. Biotechnol.* **12**, 373–385 (2011).
96. Hotta, N., Kawamori, R., Fukuda, M., Shigeta, Y. & Aldose Reductase Inhibitor-Diabetes Complications Trial Study Group. Long-term clinical effects of epalrestat, an aldose reductase inhibitor, on progression of diabetic neuropathy and other microvascular complications: multivariate epidemiological analysis based on patient background factors and severity of diabetic neuropathy. *Diabet. Med.* **29**, 1529–1533 (2012).
97. De Saedeleer, C. J. *et al.* Lactate activates HIF-1 in oxidative but not in Warburg-phenotype human tumor cells. *PLoS ONE* **7**, e46571 (2012).
98. Végran, F., Boidot, R., Michiels, C., Sonveaux, P. & Feron, O. Lactate influx through the endothelial cell monocarboxylate transporter MCT1 supports an NF- κ B/IL-8 pathway that drives tumor angiogenesis. *Cancer Res.* **71**, 2550–2560 (2011).
99. Ruan, G. X. & Kazlauskas, A. Lactate engages receptor tyrosine kinases Axl, Tie2, and vascular endothelial growth factor receptor 2 to activate phosphoinositide 3-kinase/Akt and promote angiogenesis. *J. Biol. Chem.* **288**, 21161–21172 (2013).
100. Wei, X. *et al.* *De novo* lipogenesis maintains vascular homeostasis through endothelial nitric-oxide synthase (eNOS) palmitoylation. *J. Biol. Chem.* **286**, 2933–2945 (2011).
101. Majmundar, A. J., Wong, W. J. & Simon, M. C. Hypoxia-inducible factors and the response to hypoxic stress. *Mol. Cell* **40**, 294–309 (2010).
102. Becker, J. C., Andersen, M. H., Schrama, D. & Thor Straten, P. Immune-suppressive properties of the tumor microenvironment. *Cancer Immunol. Immunother.* **62**, 1137–1148 (2013).
103. Yan, Z., Garg, S. K., Kipnis, J. & Banerjee, R. Extracellular redox modulation by regulatory T cells. *Nature Chem. Biol.* **5**, 721–723 (2009).
104. Yamanishi, S., Katsumura, K., Kobayashi, T. & Puro, D. G. Extracellular lactate as a dynamic vasoactive signal in the rat retinal microvasculature. *Am. J. Physiol. Heart Circ. Physiol.* **290**, H925–H934 (2006).
105. Ohashi, T. *et al.* Dichloroacetate improves immune dysfunction caused by tumor-secreted lactic acid and increases antitumor immunoreactivity. *Int. J. Cancer* **133**, 1107–1118 (2013).
106. Nieman, K. M. *et al.* Adipocytes promote ovarian cancer metastasis and provide energy for rapid tumor growth. *Nature Med.* **17**, 1498–1503 (2011).

Acknowledgements Supporting fellowships: FWO Fellowships (B.G., B.W.W., A.K.) and Marie Curie Fellowship (B.W.W.). Supporting grants: IUAP P7/03, long-term structural funding - Methusalem funding by the Flemish Government, FWO grants, AXA Research Fund and ERC Advanced Research Grant. We thank P. Agostinis, M. Baes, K. De Bock, S. Fendt, P. Fraisl, B. Lambrecht, A. Liston, M. Mazzone, J. Rathmell, J. Van Ginderachter, P. Verstreken and others for suggestions.

Author Contributions All authors contributed to the writing of this manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.C. (peter.carmeliet@vib-kuleuven.be).

Abnormalities in human pluripotent cells due to reprogramming mechanisms

Hong Ma^{1,2*}, Robert Morey^{3*}, Ryan C. O'Neil^{4,5}, Yupeng He^{4,5}, Brittany Daughtry^{1,2}, Matthew D. Schultz⁴, Manoj Hariharan⁴, Joseph R. Nery⁴, Rosa Castanon⁴, Karen Sabatini³, Rath D. Thiagarajan³, Masahito Tachibana^{2†}, Eunju Kang^{1,2}, Rebecca Tippner-Hedges^{1,2}, Riffat Ahmed^{1,2}, Nuria Marti Gutierrez^{1,2}, Crystal Van Dyken^{1,2}, Alim Polat^{2†}, Atsushi Sugawara², Michelle Sparman², Sumita Gokhale⁶, Paula Amato⁷, Don P. Wolf², Joseph R. Ecker^{4,8}, Louise C. Laurent³ & Shoukhrat Mitalipov^{1,2,7}

Human pluripotent stem cells hold potential for regenerative medicine, but available cell types have significant limitations. Although embryonic stem cells (ES cells) from *in vitro* fertilized embryos (IVF ES cells) represent the 'gold standard', they are allogeneic to patients. Autologous induced pluripotent stem cells (iPS cells) are prone to epigenetic and transcriptional aberrations. To determine whether such abnormalities are intrinsic to somatic cell reprogramming or secondary to the reprogramming method, genetically matched sets of human IVF ES cells, iPS cells and nuclear transfer ES cells (NT ES cells) derived by somatic cell nuclear transfer (SCNT) were subjected to genome-wide analyses. Both NT ES cells and iPS cells derived from the same somatic cells contained comparable numbers of *de novo* copy number variations. In contrast, DNA methylation and transcriptome profiles of NT ES cells corresponded closely to those of IVF ES cells, whereas iPS cells differed and retained residual DNA methylation patterns typical of parental somatic cells. Thus, human somatic cells can be faithfully reprogrammed to pluripotency by SCNT and are therefore ideal for cell replacement therapies.

The derivation of human ES cells from *in vitro* fertilized embryos¹ is relevant for cell-based therapies, and while iPS cell technology^{2,3} overcomes allogenicity issues, a high frequency of genetic and epigenetic abnormalities have been observed, including subchromosomal duplications and deletions detected as copy number variations (CNVs)^{4,5}, protein-coding mutations⁶ and defects in DNA methylation and gene expression at regions subject to imprinting and X-chromosome inactivation^{7–10}. Although it is not yet understood whether these aberrant epigenetic marks reflect errors arising during reprogramming or incomplete reversion to pluripotency, they could impact the accuracy of *in vitro* disease modelling or, more importantly, the utility of iPS cells for regenerative medicine. With the availability of somatic cell nuclear transfer as an alternative approach to somatic cell reprogramming¹¹, we explored the mechanisms underlying transcription factor- and SCNT-based reprogramming.

Genetically matched cell lines

In addition to four NT ES cell lines derived from fetal human dermal fibroblasts (HDFs), designated NT1–NT4 (ref. 11), we generated seven iPS cell lines from the same HDFs using retroviral vectors¹² (two lines, named iPS-R1 and iPS-R2) and Sendai-virus-based vectors¹³ (five lines, named iPS-S1, iPS-S2, iPS-S3, iPS-S4 and iPS-S5). Two IVF ES cell lines (human ES Oregon (hESO)-7 and hESO-8) were derived following IVF of oocytes from the same egg donor used for SCNT¹¹. All cell lines maintained typical morphology, expressed pluripotency markers, formed teratomas and retained diploid karyotypes with no detectable numerical or structural chromosomal abnormalities.

Short tandem repeat (STR) genotyping verified that all NT ES cell and iPS cell lines were genetically matched to each other and to HDFs.

The one exception to this was iPS-R1, which had a homozygous D3S1768 locus on chromosome 3 (Supplementary Table 1), whereas all other lines were heterozygous at this locus.

SNP genotyping also confirmed that all NT ES cell and iPS cell lines were essentially identical to each other and to the HDFs in terms of their nuclear genomes (>99.96% similarity, Supplementary Table 2). Oocyte and sperm donors showed first-degree genetic relationships to IVF ES cells.

Using whole methylome and transcriptome sequencing, the mitochondrial DNA (mtDNA) in NT ES cells matched those of the IVF ES cells, whereas the iPS cell and HDF sequences differed from those of the IVF ES cells at 13 nucleotide positions (Extended Data Fig. 1a, b). Consistent with previous measurements, we detected a small amount of HDF mtDNA carryover (1–4.9%) in some NT ES cells (Supplementary Table 3).

Subchromosomal aberrations

High-throughput SNP genotyping identified ten *de novo* CNVs in early-passage iPS cells and three in NT ES cells (Extended Data Fig. 2a). NT3 carried a one-copy deletion on chromosome 16, and NT4 had two duplications on chromosomes 3 and 6. Among the iPS cells, iPS-S1 harboured two duplications on chromosomes 1 and 5; iPS-S2 had three one-copy deletions on chromosomes 1, 4 and 17; iPS-S3 carried a single one-copy deletion on chromosome 10; iPS-R1 displayed two duplications on chromosomes 3 and 4, one large run of homozygosity (ROH) encompassing most of the short arm of chromosome 3 and one two-copy deletion within the ROH. This ROH was consistent with STR analysis (Supplementary Table 1). A single one-copy deletion on the X chromosome was identified in human hESO-7. All CNVs were validated using quantitative PCR (qPCR) analysis (Extended Data Table 1).

¹Center for Embryonic Cell and Gene Therapy, Oregon Health & Science University, 3303 Southwest Bond Avenue, Portland, Oregon 97239, USA. ²Division of Reproductive and Developmental Sciences, Oregon National Primate Research Center, Oregon Health & Science University, 505 Northwest 185th Avenue, Beaverton, Oregon 97006, USA. ³Department of Reproductive Medicine, University of California, San Diego, Sanford Consortium for Regenerative Medicine, 2880 Torrey Pines Scenic Drive, La Jolla, California 92037, USA. ⁴Genomic Analysis Laboratory, the Salk Institute for Biological Studies, La Jolla, California 92037, USA. ⁵Bioinformatics Program, University of California at San Diego, La Jolla, California 92093, USA. ⁶University Pathologists LLC, Boston University School of Medicine, Roger Williams Medical Center, Providence, Rhode Island 02118, USA. ⁷Division of Reproductive Endocrinology, Department of Obstetrics and Gynecology, Oregon Health & Science University, 3181 Southwest Sam Jackson Park Road, Portland, Oregon 97239, USA. ⁸Howard Hughes Medical Institute, the Salk Institute for Biological Studies, La Jolla, California 92037, USA. [†]Present addresses: Department of Obstetrics and Gynecology, South Miyagi Medical Center, Shibata-gun, Miyagi 989-1253, Japan (M.T.); Department of Cell and Molecular Biology, Karolinska Institutet, SE-17177 Stockholm, Sweden (A.P.).

*These authors contributed equally to this work.

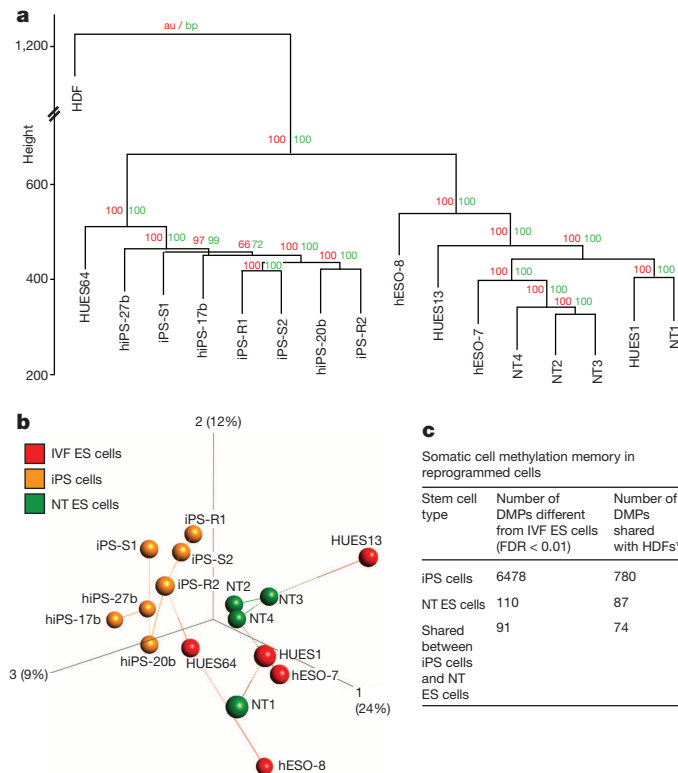


Figure 1 | Global methylation status. **a**, Unsupervised hierarchical clustering of all filtered and normalized methylation probes in five IVF ES cell lines, seven iPS cell lines, and four NT ES cell lines, and in parental HDFs. Red and green values above each edge represent AU (approximately unbiased) and BP (bootstrap probability) P values (%) calculated using bootstrap resampling¹⁶. **b**, Principal component analysis of IVF ES cells (red balls), iPS cells (orange balls), and NT ES cells (green balls) with nearest-neighbour analysis. The percentages in parentheses represent the variance explained by the respective axes. **c**, Total number of differentially methylated probes (DMPs) observed between matched iPS cells, NT ES cells and IVF ES cells ($n = 11$, Kruskal–Wallis test, FDR < 0.01). The number of DMPs shared with parental HDFs was used as a measure of the degree of somatic cell memory. *|Average β HDF – average β IVF-ES cells| > 0.3 and |average β iPS cells – average β IVF-ES cells| > 0.3.

CNV analysis was extended to a second matched set, consisting of NT ES cell (Leigh-NT1) and iPS cell lines (Leigh-iPS1, Leigh-iPS2 and Leigh-iPS3) derived from a patient with Leigh syndrome¹¹. G-banding did not reveal any numerical or chromosomal abnormalities and STR genotyping corroborated that all lines were from the Leigh patient (Leigh-fib, Supplementary Table 1). Leigh-NT1 carried oocyte mtDNA while all Leigh iPS cells inherited patient mtDNA including the homoplasmic m.8993T>G mutation¹⁴ (Extended Data Fig. 1c). Nine *de novo* CNVs were identified in this data set, including multiple CNVs in Leigh-iPS1 and Leigh-iPS3 and one each in Leigh-iPS2 and Leigh-NT1 (Extended Data Fig. 2a and Extended Data Table 1).

In summary, iPS cells, NT ES cells and IVF ES cells carried an average of 1.8, 0.8 and 0.5 CNVs per line, respectively (Extended Data Fig. 2b), with no statistically significant differences among cell types. InDel analysis using RNA sequencing (RNA-seq) also showed a trend towards fewer mutations in NT ES cells compared with iPS cells, but the differences among cell types were again not statistically significant ($P > 0.05$, Extended Data Fig. 2c, d). Thus, it seems that the mutagenic and selective pressures for both reprogramming approaches are not statistically different, but this conclusion is limited by the small numbers of cell lines analysed.

Global DNA methylation

DNA methylation is an important epigenetic mechanism contributing to cell identity, and significant differences have been reported between

iPS cells and IVF ES cells^{7,15}. Therefore, we examined genome-wide DNA methylation of our cell lines and compared them to publicly available samples using the Infinium HumanMethylation450 BeadChip. Unsupervised hierarchical clustering¹⁶ and bootstrap resampling identified two well-defined clusters, one containing all iPS cell lines and one IVF ES cell line from a previous study (HUES64 (ref. 17)), and another with all NT ES cell lines and four IVF ES cell lines (Fig. 1a, b). Intra-group variability was assessed using the coefficient of variation for each stem cell type and was found to be similar to previously reported cell lines¹⁷ (iPS cells = 0.71, NT ES cells = 0.73, IVF ES cells = 0.74; iPS cells¹⁷ = 0.73 and IVF ES cells¹⁷ = 0.72).

Comprehensive group-wise analysis revealed 6,478 differentially methylated probes (DMPs) between iPS cells and IVF ES cells (false discovery rate (FDR) < 0.01; Fig. 1c). Using the same criteria, 110 DMPs were found in NT ES cells, suggesting that NT ES cells are remarkably similar to IVF ES cells. We then asked if the DMPs could be attributed to residual epigenetic memory inherited from HDFs. Of the 6,478 DMPs in iPS cells, 780 displayed a substantial difference in DNA methylation in the same direction between iPS cells and IVF ES cells and between HDFs and IVF ES cells (average β -difference > |0.3|, where β is the ratio of intensities between methylated alleles and the sum of unmethylated and methylated alleles). Of the 110 DMPs in NT ES cells, 87 were substantially different both between NT ES cells and IVF ES cells and between HDFs and IVF ES cells (Fig. 1c). Functional enrichment analysis of probes that were highly methylated in iPS cells and HDFs compared to IVF ES cells indicated association with sequence-specific DNA binding transcription factor activity (2.02-fold enrichment, FDR < 0.0001). No significant annotation terms were found for hypermethylated probes shared by NT ES cells and HDFs. However, probes that were hypomethylated in iPS cells, NT ES cells and HDFs were enriched for loci associated with the major histocompatibility complex (MHC) class II protein complex (72-fold enrichment, FDR < 0.001).

We conclude that methylation profiles of NT ES cells are more similar to IVF ES cells than to iPS cells. Both cell types carry residual HDF epigenetic memory, but iPS cells retain eightfold more of such sites. Interestingly, nearly 80% of DMPs in NT ES cells, but only 12% in iPS cells, could be related to somatic memory, suggesting that the majority of methylation abnormalities in iPS cells result from reprogramming errors.

DNA methylation at imprinted and XCI regions

Aberrant methylation at imprinted loci has been observed in iPS cells^{7,18,19}. We interrogated previously identified imprinted regions^{7,20,21} (Fig. 2a) considering CpGs with a $\beta = 0.2$ –0.8 on the DNA methylation microarray as partially methylated, >0.8 aberrantly hypermethylated, and <0.2 hypomethylated. Our cell line variances within these imprinted regions were comparable to other independently generated cells¹⁷ (our lines, coefficient of variation = 0.27–0.36; coefficient of variation¹⁷ = 0.28–0.4). Based on unsupervised hierarchical clustering within imprinted regions, NT ES cell lines grouped closely with IVF ES cells and displayed fewer aberrantly methylated probes compared to iPS cells (Fig. 2a, b).

In terms of aberrant DNA methylation at imprinted regions, all NT ES cells displayed hypomethylation at *GNAS* (also known as *GNAS* complex locus); NT2 and NT3 were hypermethylated at probes located in the genomic region of *GNASAS* (also known as *GNAS* antisense RNA 1) and *GNAS* overlap; and NT4 was hypomethylated at *H19* (also known as imprinted maternally expressed transcript (non-protein coding)) (Fig. 2a), which corresponded with bi-allelic expression of this gene (Extended Data Table 2). All iPS cells and the hESO-7 cell line were hypermethylated at *PEG3* (also known as paternally expressed 3) (Fig. 2a), while only the iPS cells displayed hypermethylation at *MEG3* (also known as maternally expressed 3 (non-protein coding)) (Fig. 2a). These genes displayed reduced expression of corresponding transcripts (Fig. 2c; *MEG3* adjusted $P < 0.001$, average fold change, 19.8; *PEG3* adjusted $P < 0.005$, average fold change, 128.9). The *DIRAS3* (also known as *DIRAS* family, GTP-binding RAS-like 3) locus was hypermethylated in all iPS cells, but a corresponding change in gene expression was not seen (Fig. 2a).

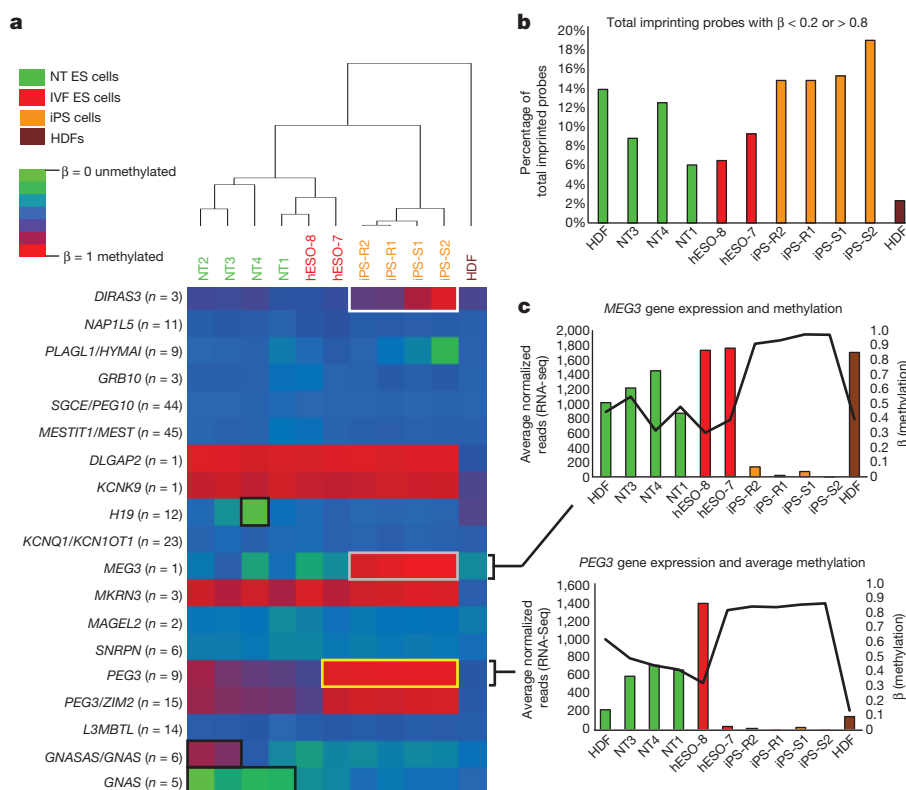


Figure 2 | Methylation at imprinted regions. **a**, Heat map of previously identified imprinted regions. For each gene, an average β -value (the ratio of intensities between methylated alleles and the sum of methylated and unmethylated alleles) for all DNA methylation probes assigned to a specific gene is shown and the number of included probes is indicated next to the gene. White box, hypermethylation at *DIRAS3* locus, no change in gene expression; black boxes, DNA methylation changes at *H19*, *GNASAS* or *GNAS*, and *GNAS*

loci (no change in gene expression); grey box, hypermethylation at the *MEG3* locus (reduced gene expression); yellow box, hypermethylation at the *PEG3* locus (reduced gene expression). **b**, Bar graph showing percentage of total imprinted probes that had a $\beta < 0.2$ or > 0.8 . **c**, Bar and line graphs showing the normalized RNA-seq read count (bars, averaged between replicates) and the DNA methylation β -values (black line) for *MEG3* and *PEG3*. Solid symbols indicate genes with overlapping genomic regions.

X-chromosome inactivation (XCI) can be detected by allele-specific expression and coating of the X chromosome by the long noncoding RNAs *XIST* (also known as X inactive specific transcript (non-protein coding)) and *XACT*^{22,23}. Based on RNA-seq, all female cells in our dataset expressed similar levels of *XIST*, but only pluripotent cells expressed *XACT* (Extended Data Fig. 3a, b). hESO-8 (male) was unmethylated at previously annotated XCI loci⁷, whereas all female lines were predominantly partially methylated ($\beta = 0.2$ – 0.8 ; Fig. 3a). NT ES cells and IVF ES cells demonstrated higher DNA methylation levels at XCI loci compared to HDFs. However, methylation levels in iPS cells were significantly higher than in NT ES cells and female hESO-7 (Fig. 3b, $P < 0.001$), with substantial variation among lines. With aberrant methylation defined as $\beta < 0.2$ or > 0.8 , NT ES cells and hESO-7 had fourfold fewer aberrations than iPS cells (Fig. 3c, $P < 0.001$). *POU3F4* (also known as POU class 3 homeobox 4), *SLITRK2* (also known as SLIT and NTRK-like family, member 2) and *SLITRK4* (also known as SLIT and NTRK-like family, member 4) hypermethylation in iPS-R2 correlated with lower gene expression while hypomethylation of *DACH2* (also known as Dachshund homologue 2), *RPS6KA6* (also known as ribosomal S6 kinase 4) and *CHM* (also known as choroideremia rab escort protein 1) in iPS-R1 and *TMEM187* (also known as transmembrane protein 187) in iPS-S2 correlated with increased gene expression (Fig. 3a and Extended Data Fig. 4a, b).

Autosomal non-imprinted loci

Differential DNA methylation analysis of autosomal non-imprinted sites revealed 1,621 DMPs between our groups (Kruskal–Wallis test, $P < 0.01$, $\Delta\beta > 0.5$). We grouped these probes into six major clusters using an unsupervised self-organizing map algorithm²⁴ (Extended Data Fig. 5). All six clusters were analysed for *cis*-regulatory functional enrichments using

GREAT²⁵, but only cluster 3 showed significant enrichments for categories associated with morphogenesis and neural development (Supplementary Table 4). iPS cells displayed higher DNA methylation levels compared to NT ES cells and IVF ES cells for most clusters, with the exception of cluster 4, in which the highest DNA methylation was seen in the IVF ES cells. NT ES cells displayed intermediate DNA methylation, but overall were closer to IVF ES cells. We examined several different probe subsets, and noted higher methylation levels in iPS cells compared to IVF ES cells, as reported previously^{26,27} (Extended Data Fig. 6a–o).

Whole-genome bisulphite sequencing

To gain a more detailed picture of the underlying methylation differences in our cells, we generated high-coverage base-resolution methylomes (14× to 25×) using MethylC-seq⁸. We also assessed the methylomes of three additional IVF ES cells (H1, H9 and HUES6)^{28–30}. Hierarchical clustering at CG differentially methylated regions (DMRs) demonstrated that the methylation landscape of NT ES cells more closely matched IVF ES cells compared to iPS cells (Fig. 4a). By comparing the methylomes and filtering regions that were obscure or highly variable in IVF ES cells, 678 CG DMRs were identified (Supplementary Table 5) that were present in at least one NT ES cell or iPS cell line but not in IVF ES cells ($FDR = 0.01$). Most of these CG DMRs were identified within iPS cells (619), whereas NT ES cells contained threefold fewer (212) and 153 CG DMRs were shared (Fig. 4b). Using a similar approach, we calculated that five previously profiled iPS cells⁸ carried a total of 792 CG DMRs, suggesting that both iPS cell groups are comparable. Most of the CG DMRs were localized within CG islands and gene bodies (Fig. 4c). Analysis of CG-DMR distribution among individual cell lines showed that each NT ES

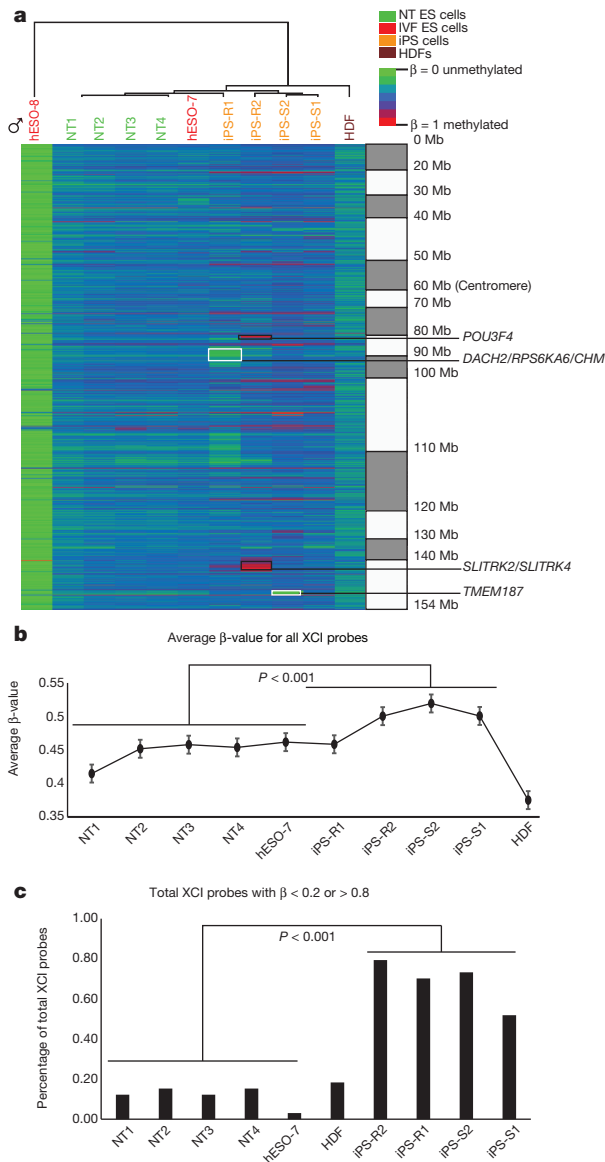


Figure 3 | Methylation at X-chromosome inactivation sites. **a**, Heat map displaying β -values of previously identified XCI probes on the DNA methylation array in NT ES cells, IVF ES cells, iPS cells and HDFs. The genes highlighted with black boxes showed both aberrant hypermethylation and corresponding changes in gene expression. The hypomethylated genes highlighted in white boxes were associated with corresponding changes in gene expression. **b**, Line graph showing an average β -value for all XCI probes for each cell line (two-sided t -test, $P < 0.001$, error bars s.e.m.). **c**, The percentage of total XCI probes with $\beta < 0.2$ or $\beta > 0.8$ (two-sided t -test, $P < 0.001$).

cell line had fewer aberrant regions than any of the iPS cell lines (Fig. 4d, $P = 0.0147$, Mann–Whitney test). CG DMRs were then assigned into three groups: memory DMRs (mDMRs; shared with HDF), NT-specific DMRs (ntDMRs) and iPS-cell-specific DMRs (iDMRs). On average, 38% of total CG DMRs in the NT ES cell lines and 22% of DMRs in iPS cells were mDMRs (Fig. 4d).

Inspection of the recurrent CG DMRs (hotspot DMRs⁸) in every iPS cell or NT ES cell line revealed that NT ES cell lines had 50 hotspot DMRs, or twofold fewer than iPS cells (104) (Fig. 4e). Interestingly, 48 of 50 hotspot DMRs in NT ES cells were also shared with iPS cells ($P < 0.001$, Hypergeometric test). Of the hotspot DMRs shared among all 8 cell lines 63% (30 out of 48) were mDMRs, suggesting regions resistant to reprogramming by either approach. Only 2 (4%) hotspot DMRs were unique to NT ES cells compared to 56 (54%) iPS-cell-specific hotspots (Fig. 4e).

Non-CG methylation in NT ES cells

We previously identified pervasive and exclusive non-CG methylation in pluripotent cells³¹. We also reported that iPS cells carry frequent aberrant non-CG methylations⁸. We identified regions showing megabase-scale non-CG methylation differences (non-CG mega DMRs) in NT ES cells and iPS cells when compared to IVF ES cells. Five IVF ES cell lines, two from this study and the three described previously, served as our control methylation landscape^{28–30}. Autosomal non-CG mega DMRs (150) were identified when the methylomes of 13 iPS cell lines and NT ES cell lines were compared to controls (Extended Data Fig. 7a and Supplementary Table 6). Non-CG mega DMRs linked to the sex chromosomes were excluded due to the mixed gender of controls. A total of 150 autosomal non-CG mega DMRs covered 123 megabases (Mb) of genome and included all regions reported previously⁸ (99% of bases); of these, 77 non-CG mega DMRs were identified from the iPS and NT cells, 70 of which occurring exclusively in iPS cells (Fig. 5a). These DMRs were distributed on every autosomal chromosome except chromosome 13 (Fig. 5b). Only 7 non-CG mega DMRs (tenfold less) were present in NT ES cells. Consistent with our previous findings⁸, non-CG mega DMRs were significantly closer to centromeric and telomeric regions compared with shuffled non-CG mega DMRs (Fig. 5b, $P < 0.001$). We also observed several different patterns of aberrant non-CG methylation, including hypomethylation in iPS cells only, or in both NT ES cells and iPS cells, and hypermethylation in iPS cells only (Extended Data Fig. 7b, c, d). However, the vast majority of non-CG mega DMRs (92.5% of total bases) were hypomethylated in iPS cells and/or NT ES cells compared with IVF ES cells (Fig. 5c).

We asked whether methylomes from our four iPS cells were similar to other iPS cells⁸. The former contained a total of 75 DMRs, while the latter carried 121, indicating that despite different somatic cell origin and culture conditions, iPS cells carried similar levels of aberrant non-CG methylation. In contrast, NT ES cells showed the least amount of aberrant non-CG methylation (Fig. 5c, d; $P < 0.005$). Hierarchical clustering for all non-CG mega DMRs also supported the conclusion that the NT ES cells are more similar to IVF ES cells (Extended Data Fig. 7a).

To understand the functional impact of non-CG mega DMRs, we focused on transcriptional activity within those regions. On average, 2 genes in NT ES cells and 30 in iPS cells were located within non-CG mega DMRs, implying that fewer genes in NT ES cells are affected (Extended Data Fig. 8a, b; $P = 0.0147$). Gene ontology analysis³² for genes in hypomethylated non-CG DMRs revealed that these genes were related to olfactory transduction, epidermal cell differentiation, cytoskeleton, immunoglobulin and homeobox proteins ($FDR \leq 0.001$; Supplementary Table 7). Gene expression in the iPS cells for 2 genes in the hypermethylated non-CG mega DMRs was upregulated (Extended Data Fig. 8c, $P < 0.05$), whereas expression of 24 genes in the iPS cells and 6 genes in the NT ES cells in the hypomethylated non-CG mega DMRs were down regulated (Extended Data Fig. 8d, e; $P < 0.001$). These observations indicate that NT ES cells were more faithfully reprogrammed to a state closely matching IVF ES cells compared to iPS cells. Particularly, NT4 had the least aberrant methylation in both CG and non-CG contexts.

Global gene expression

Lastly, we examined global gene expression patterns from strand-specific RNA-seq. Consistent with DNA methylation, intra-group variability was similar among cell types (coefficients of variation: NT ES cells = 1.41, IVF ES cells = 1.45, iPS cells = 1.44) and unsupervised hierarchical clustering positioned NT ES cells closely with IVF ES cells (Fig. 6a). Differential expression analysis ($FDR < 0.05$) yielded 1,220-transcripts, grouped into 10 clusters. The majority (65%) of these genes were either significantly upregulated or downregulated in iPS cells compared to NT ES cells and IVF ES cells. Clusters 2 and 3 showed higher gene expression in NT ES cells and IVF ES cells compared to iPS cells; when subjected to functional enrichment analysis, these clusters were associated with p38 MAPK signalling pathway genes ($FDR = 0.02$; $n = 51$) and Krueppel-associated box genes ($FDR = 0.001$; $n = 91$). Cluster 10 contained transcripts that were upregulated in

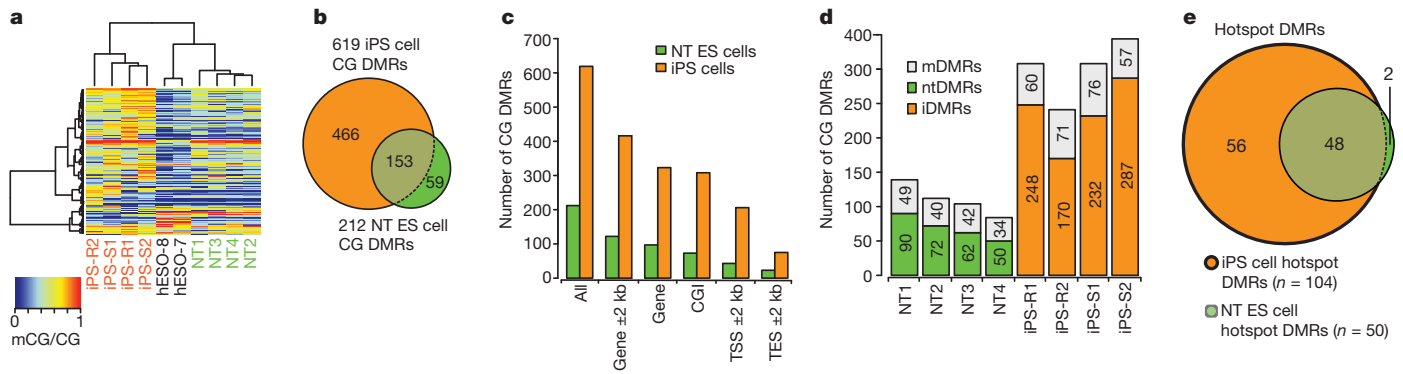


Figure 4 | CG DMRs across NT ES cells and iPS cells. **a**, Complete hierarchical clustering of CG methylation for a total 678 CG DMRs identified by comparing methylomes of NT ES cells and iPS cells to IVF ES cells. **b**, Venn diagram showing the overlap of CG DMRs across iPS cells and NT ES cells in cases in which the DMR is found in at least one of the lines in the same group. **c**, The number of 678 CG DMRs that overlapped (at least 1 bp) with indicated genomic features. CGI, CG islands; TES, transcription end sites; TSS, transcription start sites. **d**, Distribution of CG DMRs among each NT ES cell

and iPS cell line. DMRs that were also shared with parental somatic cells were identified as memory or mDMRs. Other DMRs were then assigned into NT-specific DMRs (ntDMRs) and iPS-cell-specific DMRs (iDMRs) if the DMRs were present in NT ES cell lines and iPS cell lines, respectively. **e**, The Venn diagram shows the hotspot CG DMRs that were identified in every iPS cell or NT ES cell line in the same group. Hotspot CG DMRs (48) were shared among all iPS cell and NT ES cell lines.

IVF ES cells compared to both NT ES cells and iPS cells and included genes associated with zinc finger and *C2H2*-like genes (FDR = 0.002; $n = 227$). Cluster 8 was enriched for MGI expression of TS10 primary trophoblast giant cells (FDR = 0.03; $n = 46$) and cluster 5 was associated with Y-linked inheritance.

Based on differential expression analysis, we searched for genes displaying transcriptional memory in both iPS cells and NT ES cells. Three separate *t*-tests between HDFs and IVF ES cells, NT ES cells and IVF ES cells and iPS cells and IVF ES cells were conducted at a FDR cut-off of 0.05. We found 24 genes that were expressed at significantly lower levels in the NT ES cells and HDFs compared to IVF ES cells, probably indicating incompletely reactivated genes, and 12 genes that were expressed at significantly higher levels representing incompletely silenced genes (Fig. 6b). In contrast, 171 genes were incompletely reactivated and 32 were incompletely silenced in iPS cells.

We found that incompletely reactivated genes in iPS cells also retained significantly higher promoter methylation ($P < 2.2 \times 10^{-16}$, Mann-Whitney test), possibly indicating incomplete demethylation during reprogramming (Fig. 6c). Overall, the gene expression and DNA methylation results were consistent, both suggesting that NT ES cells are more similar to IVF ES cells than to iPS cells.

Conclusions

Here, we showed that transcription-factor-based reprogramming is associated with incomplete epigenetic reprogramming. In contrast, the same somatic cells reprogrammed by SCNT displayed epigenetic and transcriptional signatures remarkably similar to those of IVF ES cell controls.

Both NT ES cells and iPS cells contained similar levels of *de novo* CNVs, with some lines (iPS-R2, iPS-S4, iPS-S5 and NT1 and NT2) displaying no detectable alterations. This observation indicates that screening of multiple cell lines may allow recovery of genetically normal lines. However, CNV analysis does not completely exclude the presence of point mutations, small indels, or translocations. Indeed, exome sequencing has demonstrated that iPS cells carry, on average, six non-synonymous point mutations per line³³.

Using genome-wide microarray-based DNA methylation as an indicator of reprogramming, we demonstrated that NT ES cells undergo more complete reprogramming than iPS cells. We also confirmed the persistence of somatic patterns of CpG methylation in human iPS cells, consistent with a mouse study³⁴. Although NT ES cells also carried evidence of epigenetic memory, iPS cells contained eightfold more CpG sites that retained the DNA methylation pattern of parental HDFs. Whole-genome bisulphite sequencing was consistent with the DNA methylation microarray analysis, showing that iPS cells carried threefold more aberrant CG and tenfold more aberrant non-CG methylation compared to NT ES cells, indicating that SCNT reprogramming is capable of resetting the DNA methylation and corresponding gene expression program more faithfully than iPS cell reprogramming.

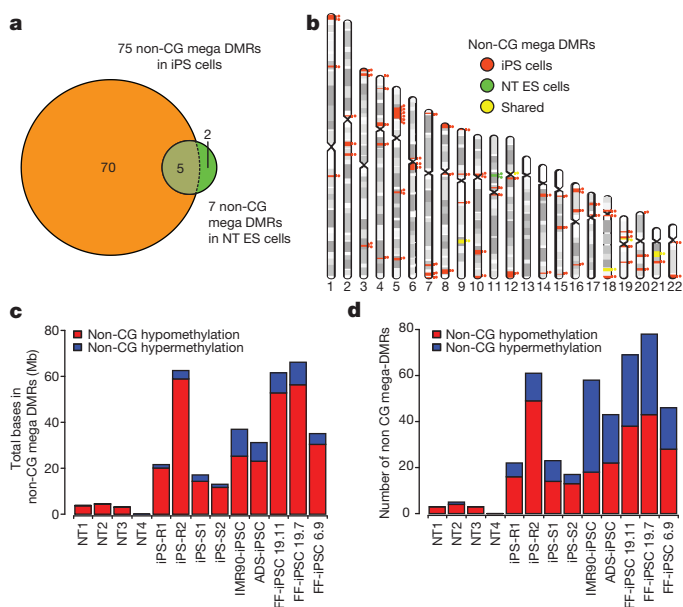


Figure 5 | Non-CG mega DMRs in NT ES cells and iPS cells. **a**, Venn diagram showing the overlap of the 77 non-CG mega DMRs identified in the iPS cell and the NT ES cell lines from this study. Numbers within circles denote DMRs identified exclusively within each group. Five DMRs were shared among all cell lines in both groups. **b**, Chromosome ideogram showing the location of the 77 non-CG mega DMRs found in both NT ES cell and iPS cell lines from this study. Orange circles and lines indicate the location of the individual DMRs specific for iPS cells; green circles and lines denote those specific for NT ES cells and yellow circles and lines are DMRs shared by both cell types. **c**, Total length of the non-CG mega DMRs identified in 4 NT ES cell and 9 iPS cell lines. The NT ES cells had a significantly lower size of DMRs (Mann-Whitney test, $P < 0.005$) compared to the iPS cells. FF, foreskin fibroblasts. **d**, Total number of the non-CG mega DMRs identified in the cell lines. The NT ES cells had a significantly lower number of DMRs (Mann-Whitney test, $P < 0.005$) compared to the iPS cells.

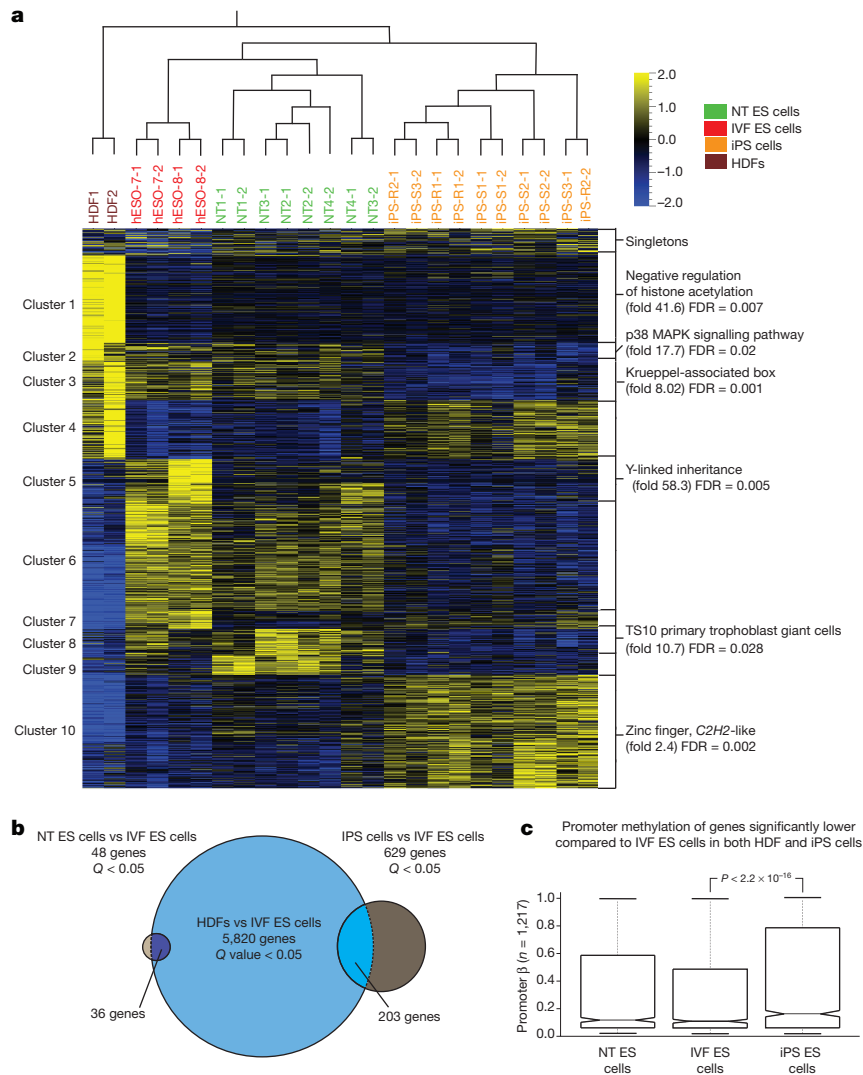


Figure 6 | Gene expression analysis by RNA-seq. **a**, Heat map displaying 1,220 differentially expressed genes between NT ES cells, iPS cells and IVF ES cells ($n = 22$) (ANOVA adjusted p -value < 0.05). Genes were clustered into ten-groups for functional analysis and presented as a heat map. Cluster 4, 6, 7, and 9 showed no significant functional enrichments. **b**, Venn diagram showing the number of genes differentially expressed between the HDFs and the IVF ES cells (large circle), the iPS cells and the IVF ES cells (medium circle) and the NT ES cells and IVF ES cells (small circle; t -test FDR < 0.05). Overlapping regions represent the number of genes differentially expressed in both the

HDFs and either the NT ES cells or iPS cells. **c**, Notched box plots represent the β -value of all probes in the promoter regions ($-2,000$ bp to 500 bp) of the genes that were expressed at significantly lower levels (t -test FDR < 0.05) in both the HDFs and the iPS cells (exhibiting transcriptional memory) when compared to the IVF ES cells. The box represents the interquartile range (25th to 75th percentile), and the line within the box marks the median. The notch in the box represents the 95% confidence interval around the median. The whiskers above and below the box contain 99.3% of the data and the number of CpGs interrogated is shown on the y axis.

An explanation for this more effective reprogramming by SCNT is that the ooplasm provides 'physiologic' levels of reprogramming factors that are upstream of pluripotency. It has been suggested that oocyte factors rapidly demethylate the somatic genome, whereas this process occurs passively during factor-based reprogramming³⁴. Clearly, elucidation of oocyte-based reprogramming mechanisms will support the development of improved reprogramming protocols.

In summary, although IVF ES cells most closely resemble cells residing in embryos, they are allogeneic. Human iPS cells might remain the most facile cell type for many *in vitro* applications, but show extensive epigenetic and transcriptomic aberrations compared to NT ES cells and IVF ES cells. NT ES cells combine significant advantages of both types; epigenetic stability of IVF ES cells and the histocompatible nature of iPS cells. Further studies on additional NT ES cell lines, especially lines derived from aged patients, and their differentiation potential are now warranted.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 December 2013; accepted 5 June 2014.

Published online 2 July 2014.

- Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
- Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
- Rais, Y. *et al.* Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502**, 65–70 (2013).
- Hussein, S. M. *et al.* Copy number variation and selection during reprogramming to pluripotency. *Nature* **471**, 58–62 (2011).
- Laurent, L. C. *et al.* Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPS cells during reprogramming and time in culture. *Cell Stem Cell* **8**, 106–118 (2011).
- Ruiz, S. *et al.* Analysis of protein-coding mutations in hiPSCs and their possible role during somatic cell reprogramming. *Nature Commun.* **4**, 1382 (2013).

7. Nazor, K. L. *et al.* Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. *Cell Stem Cell* **10**, 620–634 (2012).
8. Lister, R. *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471**, 68–73 (2011).
9. Ohi, Y. *et al.* Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. *Nature Cell Biol.* **13**, 541–549 (2011).
10. Ruiz, S. *et al.* Identification of a specific reprogramming-associated epigenetic signature in human induced pluripotent stem cells. *Proc. Natl Acad. Sci. USA* **109**, 16196–16201 (2012).
11. Tachibana, M. *et al.* Human embryonic stem cells derived by somatic cell nuclear transfer. *Cell* **153**, 1228–1238 (2013).
12. Lowry, W. E. *et al.* Generation of human induced pluripotent stem cells from dermal fibroblasts. *Proc. Natl Acad. Sci. USA* **105**, 2883–2888 (2008).
13. Fusaki, N., Ban, H., Nishiyama, A., Saeki, K. & Hasegawa, M. Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proc. Jpn Acad. B* **85**, 348–362 (2009).
14. Taylor, R. W. & Turnbull, D. M. Mitochondrial DNA mutations in human disease. *Nature Rev. Genet.* **6**, 389–402 (2005).
15. Bock, C. *et al.* Reference maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* **144**, 439–452 (2011).
16. Suzuki, R. & Shimodaira, H. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006).
17. Ziller, M. J. *et al.* Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. *PLoS Genet.* **7**, e1002389 (2011).
18. Stelzer, Y. *et al.* Identification of novel imprinted differentially methylated regions by global analysis of human-parthenogenetic-induced pluripotent stem cells. *Stem Cell Rep.* **1**, 79–89 (2013).
19. Rugg-Gunn, P. J., Ferguson-Smith, A. C. & Pedersen, R. A. Status of genomic imprinting in human embryonic stem cells as revealed by a large cohort of independently derived and maintained lines. *Hum. Mol. Genet.* **16**, R243–R251 (2007).
20. de Hoon, M. J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20**, 1453–1454 (2004).
21. Saldanha, A. J. Java Treeview—extensible visualization of microarray data. *Bioinformatics* **20**, 3246–3248 (2004).
22. Silva, S. S., Rowntree, R. K., Mekhoubad, S. & Lee, J. T. X-chromosome inactivation and epigenetic fluidity in human embryonic stem cells. *Proc. Natl Acad. Sci. USA* **105**, 4820–4825 (2008).
23. Vallot, C. *et al.* XACT, a long noncoding transcript coating the active X chromosome in human pluripotent cells. *Nature Genet.* **45**, 239–241 (2013).
24. Newman, A. M. & Cooper, J. B. AutoSOME: a clustering method for identifying gene expression modules without prior knowledge of cluster number. *BMC Bioinformatics* **11**, 117 (2010).
25. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnol.* **28**, 495–501 (2010).
26. Nishino, K. *et al.* DNA methylation dynamics in human induced pluripotent stem cells over time. *PLoS Genet.* **7**, e1002085 (2011).
27. Polo, J. M. *et al.* Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nature Biotechnol.* **28**, 848–855 (2010).
28. Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148 (2013).
29. Laurent, L. *et al.* Dynamic changes in the human methylome during differentiation. *Genome Res.* **20**, 320–331 (2010).
30. Lister, R. *et al.* Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013).
31. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
32. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.* **25**, 25–29 (2000).
33. Gore, A. *et al.* Somatic coding mutations in human induced pluripotent stem cells. *Nature* **471**, 63–67 (2011).
34. Kim, K. *et al.* Epigenetic memory in induced pluripotent stem cells. *Nature* **467**, 285–290 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors acknowledge the OHSU Embryonic Stem Cell Research Oversight Committee and the Institutional Review Board for providing oversight and guidance. We thank oocyte and sperm donors and the Women's Health Research Unit staff at the Center for Women's Health, University Fertility Consultants and the Reproductive Endocrinology and Infertility Division in the Department of Obstetrics and Gynecology of Oregon Health and Science University for their support and procurement of human gametes. We are grateful to C. Penedo for microsatellite analysis and W. Sanger and D. Zaleski for karyotyping services. We are also indebted to Y. Li, H. Sritanandomchai and D. Melguizo Sanchis for their technical support. We thank the staff at the Institute for Genomic Medicine Genomics Facility at UCSD for running the Infinium HumanMethylation450 BeadChips and sequencing of the RNA-seq libraries. The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin (<http://www.tacc.utexas.edu>) and the San Diego Supercomputing Center (through an allocation from the eXtreme Science and Engineering Discovery Environment (XSEDE)) for providing HPC resources that have contributed to the research results reported within this paper. SCNT and iPS cell studies were supported by grants from the Leduca Foundation and OHSU institutional funds. R.M., K.S., R.T. and L.C.L. were supported by the UCSD Department of Reproductive Medicine. Methylome studies were supported by the Salk International Council Chair fund endowment and the Mary K. Chapman Foundation to J.R.E. J.R.E. is an investigator of the Howard Hughes Medical Institute and the Gordon and Betty Moore Foundation (GMBF3034). A.P. received a fellowship from the Swedish Research Council, Vetenskapsrådet. E.K. was partially funded by a fellowship from the Collins Medical Trust.

Author Contributions H.M., R.M., L.C.L. and S.M. conceived the study and designed the experiments. P.A., M.S. and N.M.G. coordinated recruitment of gamete donors. P.A. performed ovarian stimulations and oocyte retrievals. M.T., M.S., N.M.G. and S.M. conducted SCNT, IVF and embryo culture experiments. R.T.-H., S.M., M.T., M.S., N.M.G., H.M., A.P., B.D., E.K., A.S. and R.A. derived and cultured IVF ES cells, NT ES cells and iPS cells. S.G. performed teratoma analysis. H.M., M.T. and C.V.D. performed the DNA and RNA extractions, mtDNA amplification refractory mutation system qPCR analyses, and qPCR. R.M., K.S., R.D.T. and L.C.L. performed SNP, DNA methylation and RNA-seq studies and bioinformatic analysis of the data. R.C.O., Y.H., M.D.S., M.H., J.R.N., R.C. and J.R.E. conducted MethylC-seq studies. H.M., R.M., R.C.O., Y.H., J.R.E., L.C.L., D.P.W. and S.M. wrote the paper.

Author Information Processed data sets can be downloaded from the NCBI GEO under accession GSE53096 for RNA-seq, SNP array and 450K methylation array, and accession GSE57179 for MethylC-seq data. Analysed MethylC-seq data sets can also be accessed at <http://neomorph.salk.edu/SCNT/browser.html>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.M. (mitalipo@ohsu.edu), L.C.L. (lilaurent@ucsd.edu) or J.R.E. (ecker@salk.edu).

Aryl hydrocarbon receptor control of a disease tolerance defence pathway

Alban Bessede^{1,2*}, Marco Gargaro^{1*}, Maria T. Pallotta¹, Davide Matino¹, Giuseppe Servillo¹, Cinzia Brunacci¹, Silvio Biciatto³, Emilia M. C. Mazza³, Antonio Macchiarulo⁴, Carmine Vacca¹, Rossana Iannitti¹, Luciana Tissi¹, Claudia Volpi¹, Maria L. Belladonna¹, Ciriana Orabona¹, Roberta Bianchi¹, Tobias V. Lanz^{5,6}, Michael Platten^{5,6}, Maria A. Della Fazio¹, Danilo Piobbico¹, Teresa Zelante¹, Hiroshi Funakoshi⁷, Toshikazu Nakamura⁸, David Gilot⁹, Michael S. Denison¹⁰, Gilles J. Guillemin¹¹, James B. DuHadaway¹², George C. Prendergast¹², Richard Metz¹³, Michel Geffard², Louis Boon¹⁴, Matteo Pirro¹⁵, Alfonso Iorio¹⁶, Bernard Veyret², Luigina Romani¹, Ursula Grohmann¹, Francesca Fallarino¹ & Paolo Puccetti¹

Disease tolerance is the ability of the host to reduce the effect of infection on host fitness. Analysis of disease tolerance pathways could provide new approaches for treating infections and other inflammatory diseases. Typically, an initial exposure to bacterial lipopolysaccharide (LPS) induces a state of refractoriness to further LPS challenge (endotoxin tolerance). We found that a first exposure of mice to LPS activated the ligand-operated transcription factor aryl hydrocarbon receptor (AhR) and the hepatic enzyme tryptophan 2,3-dioxygenase, which provided an activating ligand to the former, to down-regulate early inflammatory gene expression. However, on LPS rechallenge, AhR engaged in long-term regulation of systemic inflammation only in the presence of indoleamine 2,3-dioxygenase 1 (IDO1). AhR-complex-associated Src kinase activity promoted IDO1 phosphorylation and signalling ability. The resulting endotoxin-tolerant state was found to protect mice against immunopathology in Gram-negative and Gram-positive infections, pointing to a role for AhR in contributing to host fitness.

Although lipopolysaccharide (LPS)-induced proinflammatory molecules are indispensable for counteracting the growth and dissemination of Gram-negative bacteria, overproduction can lead to sepsis syndrome, also known as endotoxin shock. However, a prior exposure to a low level of LPS induces a durable state of cell refractoriness to subsequent LPS challenge, a condition known as endotoxin tolerance^{1–4}. The complex events underlying this phenomenon remain poorly understood, despite a recent resurgence of interest in this effect, which involves multiple downstream effector cells and mechanisms. Endotoxin tolerance, not merely amounting to a shutdown of LPS-induced responses, involves reprogramming of gene expression⁵ and chromatin remodelling⁶. The occurrence of endotoxin tolerance has been reported in several disease settings, including sepsis, trauma, surgery, and pancreatitis, underlining its clinical significance. Pharmacologic modulation of LPS-responsive genes to accelerate the onset of endotoxin tolerance would be beneficial in clinical settings dominated by acute hyperinflammatory responses to infection⁷. In addition, endotoxin derived from Gram-negative bacteria is known to have immunomodulatory and allergy-protective potential⁷.

In experimental models, the aryl hydrocarbon receptor (AhR), a ligand-operated transcription factor^{8,9}, participates in the transcriptional regulation of several LPS-responsive genes, and AhR-deficient mice are more sensitive to endotoxin shock than wild-type mice¹⁰. This suggests a crucial function of AhR in modulating the inflammatory response mediated by LPS and Toll-like receptor (TLR)4 signalling⁵. Yet, the nature of the endogenous ligands that drive AhR-regulated gene expression in response to TLR4 activation has been unclear¹¹. Also, the

potential role of AhR in endotoxin tolerance has never been addressed experimentally.

The first step in tryptophan catabolism is the cleavage of the 2,3-double bond of the indole ring of tryptophan¹². In mammals, this reaction is performed independently by indoleamine 2,3-dioxygenase 1 (IDO1), tryptophan 2,3-dioxygenase (TDO or TDO2; mostly expressed in the liver), and the recently discovered indoleamine 2,3-dioxygenase 2 (IDO2; a paralogue of IDO1). When induced by proinflammatory cytokines¹³, tryptophan degradation by IDO1 yields a series of catabolites—collectively known as kynurenines¹⁴—that regulate immune homeostasis by acting as AhR ligands and allowing the generation of regulatory T cells^{15–17}, which protect mice from chronic hyperinflammatory responses¹⁸. Increased plasma kynurenine levels and kynurenine-to-tryptophan ratios have been found in patients with systemic inflammatory response syndrome, sepsis and septic shock, but the biological significance and prognostic value of these findings have remained uncertain¹⁹.

In the current study we used C57BL/6 wild-type mice, AhR-deficient mice, and mice lacking IDO1, IDO2 or TDO2 to investigate the role of AhR and tryptophan catabolism in primary LPS responsiveness and in the induction of endotoxin tolerance. We found that overreacting responses to primary LPS challenge were mitigated by AhR and TDO2-dependent tryptophan catabolism. Endotoxin tolerance, in contrast, required the combined effects of AhR, IDO1 and the cytokine transforming growth factor- β (TGF- β). The protective, LPS-triggered tolerant state was not restricted to immunopathology induced by LPS or

¹Department of Experimental Medicine, University of Perugia, 06132 Perugia, Italy. ²IMS Laboratory, University of Bordeaux, 33607 Pessac, France. ³Center for Genome Research, University of Modena and Reggio Emilia, 41125 Modena, Italy. ⁴Department of Chemistry and Technology of Drugs, University of Perugia, 06123 Perugia, Italy. ⁵Experimental Neuroimmunology Unit, German Cancer Research Center, 69120 Heidelberg, Germany. ⁶Department of Neurooncology, University Hospital, 69120 Heidelberg, Germany. ⁷Center for Advanced Research and Education, Asahikawa Medical University, 078-8510 Asahikawa, Japan. ⁸Kringle Pharma Joint Research Division for Regenerative Drug Discovery, Center for Advanced Science and Innovation, Osaka University, 565-0871 Osaka, Japan. ⁹CNRS UMR6290, Institut de Génétique et Développement de Rennes, Université de Rennes 1, 35043 Rennes, France. ¹⁰Department of Environmental Toxicology, University of California, Davis, 95616 California, USA. ¹¹Australian School of Advanced Medicine (ASAM), Macquarie University, 2109 New South Wales, Australia. ¹²Lankenau Institute for Medical Research, Wynnewood, 19096 Pennsylvania, USA. ¹³New Link Genetics Corporation, Ames, 50010 Iowa, USA. ¹⁴Bioceros, 3584 Utrecht, The Netherlands. ¹⁵Department of Medicine, University of Perugia, 06132 Perugia, Italy. ¹⁶Department of Clinical Epidemiology & Biostatistics, McMaster University, Ontario L8S 4K1, Canada.

*These authors contributed equally to this work.

Gram-negative bacteria, in that it also specifically targeted inflammatory cytokine production in a *Streptococcus*-induced multifocal septic arthritis model.

Endotoxaemia in *Ahr*^{-/-} and *Tdo2*^{-/-} mice

During sepsis, activation of phagocytes leads to tissue damage by migratory neutrophils and overproduction of proinflammatory cytokines, which worsen systemic inflammation and result in haemodynamic changes, multiple organ failure and ultimately death. The synthesis of acute phase proteins occurs in the liver. We investigated the effects of a first, sublethal dose (10 mg per kg, intraperitoneally) of LPS in C57BL/6 wild-type, *Ido1*^{-/-}, *Ido2*^{-/-}, *Tdo2*^{-/-} and *Ahr*^{-/-} mice (Fig. 1a). Susceptibility to endotoxaemia was greatly increased in AhR knockout (KO) and TDO2 KO mice with respect to wild-type controls, with limited (IDO1) or no (IDO2) effects of either indoleamine 2,3-dioxygenase deficiency. LPS induced an increase in serum kynurenine-to-tryptophan (Kyn/Trp) ratios, an effect that was selectively negated by *Tdo2* deficiency (Fig. 1b).

Mice of the different genotypes were challenged with a wide range (2.5–80 mg per kg) of LPS doses, to calculate the median lethal dose (LD₅₀) of endotoxin (Fig. 1c). LD₅₀ values (mg per kg) varied among

the different genotypes, being highest in wild-type (27.3), *Ido1*^{-/-} (22.5) and *Ido2*^{-/-} (32.1) mice, and considerably lower in *Tdo2*^{-/-} (9.7) and *Ahr*^{-/-} (5.2) mice. Consistent results were obtained using either the IDO1 and IDO2 inhibitor 1-methyltryptophan (1-MT) or the TDO2-selective inhibitor 680C91 (Extended Data Fig. 1a). LD₅₀ values varied between inhibitor-treated mice, being considerably higher in 1-MT-treated (22.8) than 680C91-treated (7.2) mice (Extended Data Fig. 1b). Histopathology of the lungs (Fig. 1d) and liver (Fig. 1e) revealed that inflammation and neutrophil infiltration were most intense in AhR KO and TDO2 KO mice. This suggested that, upon primary LPS challenge, IDO1 and IDO2 may not contribute significantly to protection²⁰ and that AhR and hepatic TDO2 are, in contrast, functionally important for protection.

Analysis of blood samples from wild-type, *Ahr*^{-/-}, *Ido1*^{-/-} or *Tdo2*^{-/-} mice at different times of LPS administration (10 mg per kg) showed a marked increase in plasma concentrations of TNF- α and IL-6 (peaking at 2 h under all conditions, and only marginally declining thereafter in *Ahr*^{-/-} and *Tdo2*^{-/-} mice), IL-1 β (peak levels at 6 h, further increasing at 24 h in *Ahr*^{-/-} and *Tdo2*^{-/-} mice), and IL-10 (with high, sustained production only in wild-type and *Ido1*^{-/-} mice; Fig. 1f). Thus, in accordance with previous data in the literature, resistance to primary LPS challenge correlated with an early (2–6 h), yet reversible (24 h), increase in plasma levels of TNF- α , IL-6 and IL-1 β —all dysregulated by AhR deficiency^{10,21,22}—and with a progressive increase in IL-10 (in wild-type and *Ido1*^{-/-} mice, as opposed to the stable levels in *Ahr*^{-/-} and *Tdo2*^{-/-} mice). The AhR- and TDO2-dependent induction of IL-10 was crucial in regulating the proinflammatory response to endotoxin in wild-type mice (Extended Data Fig. 2a,b). Exogenous IL-10 (250 ng per mouse, daily, administered from challenge through day 5) significantly improved survival in both TDO2- and AhR-deficient mice receiving 10 mg per kg LPS on day 0 (Extended Data Fig. 2c). TGF- β was not detectable under any conditions over the timeframe of observation.

Real-time PCR analysis revealed that, in response to 10 mg per kg LPS, AhR was expressed early (24 h) at the injection site, in peritoneal exudate cells (PECs). No modulation was observed in the liver. *Tdo2* was induced in the liver at 6–24 h of LPS injection, whereas no induction was observed in PECs (Fig. 1g). Also, no induction at all of *Ahr* and *Tdo2* occurred in *Thr4*^{-/-} mice. Protein expressions of AhR, TDO2, IDO1 and IDO2 were assessed at 24 h by immunoblot analysis in PECs and liver cells, demonstrating a positive correlation with the PCR data (Fig. 1h). Concurrent with the hepatic TDO2 expression was an increase in serum L-kynurenine-to-tryptophan ratios, an effect that was blocked by the TDO2 inhibitor 680C91, but not by *Ido1* deficiency (Fig. 1i).

To ascertain whether AhR expression correlated with functional activity, we measured induction of AhR-mediated gene transcription *ex vivo* in cells from wild-type mice, treated with 10 mg per kg LPS. Real-time PCR was used to assess transcription of a gene (*Cyp1a1*) whose promoter contains xenobiotic response elements (Fig. 1j). *Cyp1a1* was induced in PECs and liver cells at 24 h, and this induction was dependent on TDO2 (and not on IDO1) and was restored in TDO2 inhibitor-treated mice by L-kynurenine, the first by-product of tryptophan catabolism in the kynurenine pathway and an AhR signalling pathway inducer¹⁶.

L-kynurenine as an endogenous AhR ligand

The nature of L-kynurenine as an AhR ligand²³ was preliminarily investigated in hepatocytes (which abundantly express the receptor) by binding competition experiments with [³H]2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD; Fig. 2a). Molecular modelling and docking studies, and dynamics and free-energy calculations entered into a homology model of the PAS-B domain of AhR supported the ability of L-kynurenine to interact with the receptor, identifying Gln 377 as a key residue involved in the effective binding of the ligand (that is, through hydrogen bonds in the best-scoring energetic binding model). This residue was not previously reported as being part of the fingerprint residues necessary for the binding of TCDD to AhR²⁴ (Fig. 2b, c).

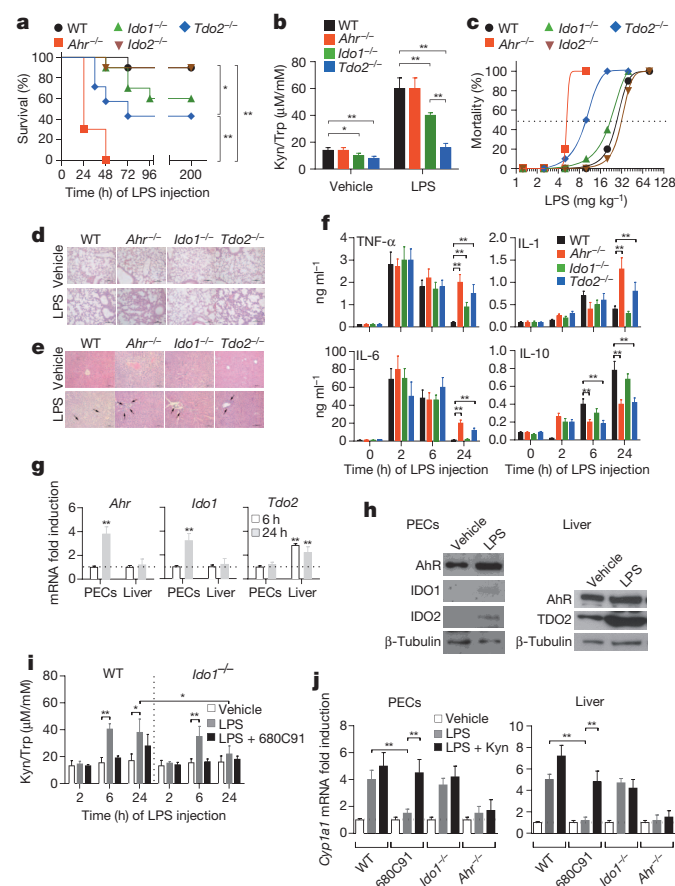


Figure 1 | Increased susceptibility of *Tdo2*^{-/-} and *Ahr*^{-/-} mice to primary LPS challenge. **a**, Survival data of challenged mice ($n = 10$). One of three experiments. * $P < 0.05$; ** $P < 0.001$ (log-rank test). **b**, Kyn/Trp ratios. Mean \pm s.d. of three experiments. * $P < 0.05$; ** $P < 0.001$ (two-tailed Student's *t*-test). **c**, Estimation of LD₅₀ ($n = 10$). **d**, **e**, Histopathology in lungs and liver, respectively. Scale bar, 100 μ m. **f**, Cytokine measurements (means \pm s.d.) from three experiments ($n = 6$; ** $P < 0.001$; two-tailed Student's *t*-test). **g**, Gene transcript expressions. Means \pm s.d. (three experiments; ** $P < 0.001$; Shapiro test). PECs, peritoneal exudate cells. **h**, Immunoblotting data. One experiment of three. **i**, Kyn/Trp ratios (means \pm s.d. of three experiments). * $P < 0.05$; ** $P < 0.001$ (two-tailed Student's *t*-test). **j**, *Cyp1a1* transcript expression; three experiments; ** $P < 0.001$ (Shapiro test).

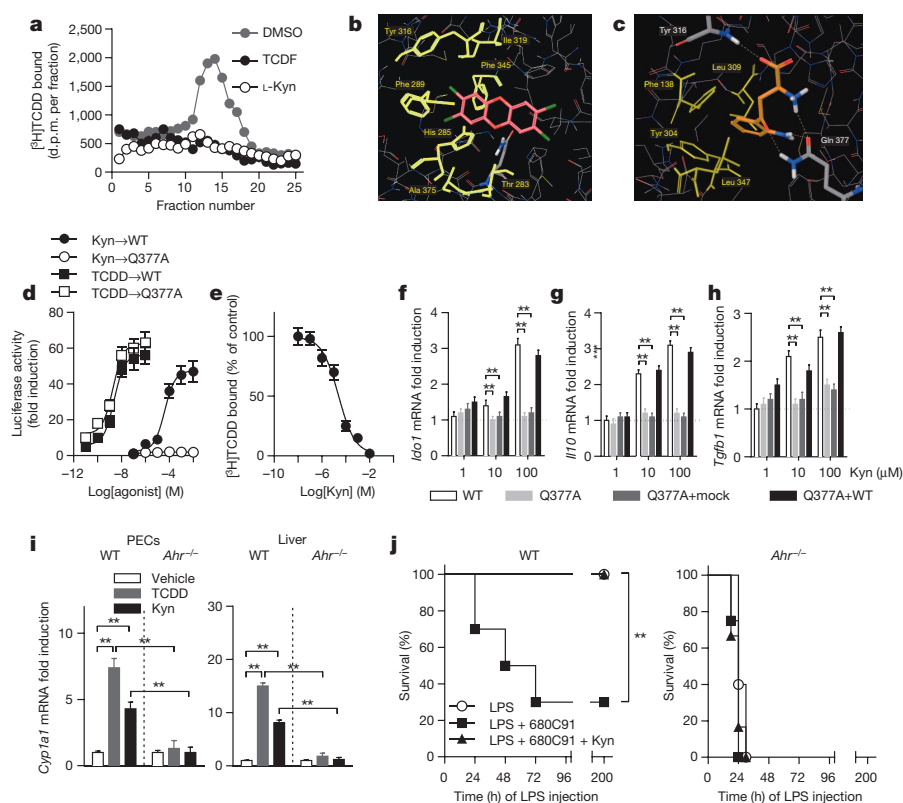


Figure 2 | L-kynurenine (Kyn) is an endogenous AhR ligand. **a**, Kyn competes with [3 H]TCDD for binding cytosolic AhR. **b**, Proposed binding mode of TCDD into the homology model of PAS-B of AhR. **c**, Proposed binding mode of Kyn (orange carbon atoms) into the homology model of PAS-B of AhR. Hydrophobic residues involved in the binding of Kyn are shown as yellow sticks. Gln 377 is shown in sticks with carbon atoms in grey. **d**, Transactivation activity of wild-type (WT) or Q377A AhR by Kyn or TCDD. **e**, Competition by Kyn for specific binding of [3 H]TCDD to cytosolic AhR from reconstituted *Ahr*^{-/-} DCs, allowing for an estimated Kyn IC₅₀ value of 36.2 μ M and of 21.6 μ M for K_i, $n = 3$; means \pm s.d. in one experiment of three. **f–h**, *Idol* (**f**), *Il10* (**g**) and *Tgfb1* (**h**) transcripts. Means \pm s.d. of three experiments; ** $P < 0.001$ (Shapiro test). **i**, *Cyp1a1* transcripts. Compiled data (means \pm s.d.) from three experiments. ** $P < 0.001$ (Shapiro test). **j**, Survival curves of variously treated mice ($n = 10$; one of three experiments). ** $P < 0.001$ (log-rank test).

To substantiate the *in silico* data, a mutant AhR receptor was engineered carrying a Gln377Ala mutation (Q377A), and AhR-deficient conventional dendritic cells (cDCs, splenic CD11c⁺ CD11b⁺ DCs) were reconstituted with wild-type or mutated AhR. Transfection efficiency (about 50%) was similar for the two constructs, as revealed by staining with antibody recognizing both wild-type and mutant AhR on fluorescence-activated cell sorting analysis. No differences in protein half-life (≈ 8 h) were found in cDCs reconstituted with either wild-type or AhR(Q377A) (Extended Data Fig. 3a, b). When assayed for *Cyp1a1* transcriptional activity in cDCs reconstituted with mutated AhR, L-kynurenine exhibited complete loss of biologic activity, in terms of potency and likely affinity, compared to wild-type receptor-transfected counterparts (Fig. 2d). In contrast, on stimulation with TCDD, the 50% maximal effective concentration (EC₅₀) value of TCDD was approximately 20-fold lower when cells had been reconstituted with the mutated AhR (Fig. 2d), pointing to greater potency and affinity of the ligand for AhR(Q377A). Competition by L-kynurenine for specific binding of [3 H]TCDD to cytosolic AhR from reconstituted *Ahr*^{-/-} DCs allowed for an estimated L-kynurenine IC₅₀ value of 36.2 μ M and of 21.6 μ M for K_i (Fig. 2e). Thus L-kynurenine and TCDD bind AhR with distinct affinities and exploit different types of interactions with AhR, either interaction involving a unique set of fingerprints residues necessary for effective binding.

AhR-deficient cDCs reconstituted with wild-type AhR, but not AhR (Q377A), transcriptionally expressed *Idol* (Fig. 2f), *Il10* (Fig. 2g) and *Tgfb1* (Fig. 2h) in a dose-dependent manner in response to L-kynurenine, further underlining the crucial requirement for Gln 377 in the AhR-dependent transcriptional activity in our setting. Cotransfection of Q377A and wild-type AhR (but not a mock control) resulted in *Idol*, *Il10* and *Tgfb1* transcription (Fig. 2f–h), indicating that the mutant AhR did not interfere with wild-type AhR function. Altogether, these data supported the binding mode of L-kynurenine to AhR as predicted by docking studies, and they explained the AhR requirement for kynurenine induction of *Cyp1a1* transcription in PECs and liver cells (Fig. 2i).

When 200 mg per kg L-kynurenine (first-order elimination kinetics, $K_E = 0.023 \text{ min}^{-1}$; resulting in physiologically attainable concentrations

in the low micromolar range) was administered to mice receiving 10 mg per kg LPS and the TDO2 inhibitor, all mice survived challenge, and this effect was negated by AhR deficiency (Fig. 2j). Overall, these data confirmed the crucial role of TDO2-dependent tryptophan catabolism in the activation of AhR, which is, in turn, necessary for downregulating inflammatory gene expression in primary endotoxaemia. These data were also consistent with previous data in a different setting showing that L-kynurenine requires AhR in the absence of functional TDO2²⁵.

IDO1, AhR and TGF- β in LPS tolerance

Next we examined the effect of high-dose LPS challenge in wild-type, IDO1-deficient or IDO2-deficient mice surviving a primary dose of 10 mg per kg LPS administered 1 week earlier. Primed mice, as well as sham-tolerized controls of the different genotypes, received 40 mg per kg LPS intraperitoneally, to be monitored for mortality parameters (Fig. 3a). Only LPS-primed wild-type mice and *Ido2*^{-/-} mice survived rechallenge beyond the 30-day observation period ($P < 0.0001$ vs *Ido1*^{-/-} and unprimed wild-type mice), as opposed to the rapid onset of a lethal shock in the other groups, whose median survival times were very similar (that is, 24–48 h). Notably, under those conditions of primary LPS challenge (10 mg per kg), the LD₅₀ of LPS-primed (that is, tolerant) wild-type mice consistently exceeded 80 mg per kg, when estimated 1 week after a first exposure to LPS. The LD₅₀ values of primed *Ido1*^{-/-} and unprimed wild-type mice were similar (27.1 and 33.9 mg per kg, respectively; Fig. 3b). Histopathology of lungs and liver showed that, at 48 h of rechallenge with 40 mg per kg LPS, inflammation and neutrophil infiltration were similar in LPS-primed *Ido1*^{-/-} mice and in IDO1-competent, unprimed controls (Fig. 3c, d).

After LPS rechallenge, increased transcriptional expression of *Idol* and increased protein levels were demonstrable in splenic CD11c⁺ CD11b⁺ DCs (usually referred to as conventional DCs, or cDCs) and only marginally in peritoneal macrophages, and not at all in plasmacytoid DCs (pDCs) and peritoneal neutrophils (Extended Data Fig. 4a, b). LPS-primed wild-type and *Ido1*^{-/-} cDCs promptly upregulated *Ahr* transcription upon LPS restimulation, yet this translated into enhanced AhR-dependent transcriptional activity only in IDO1-competent DCs,

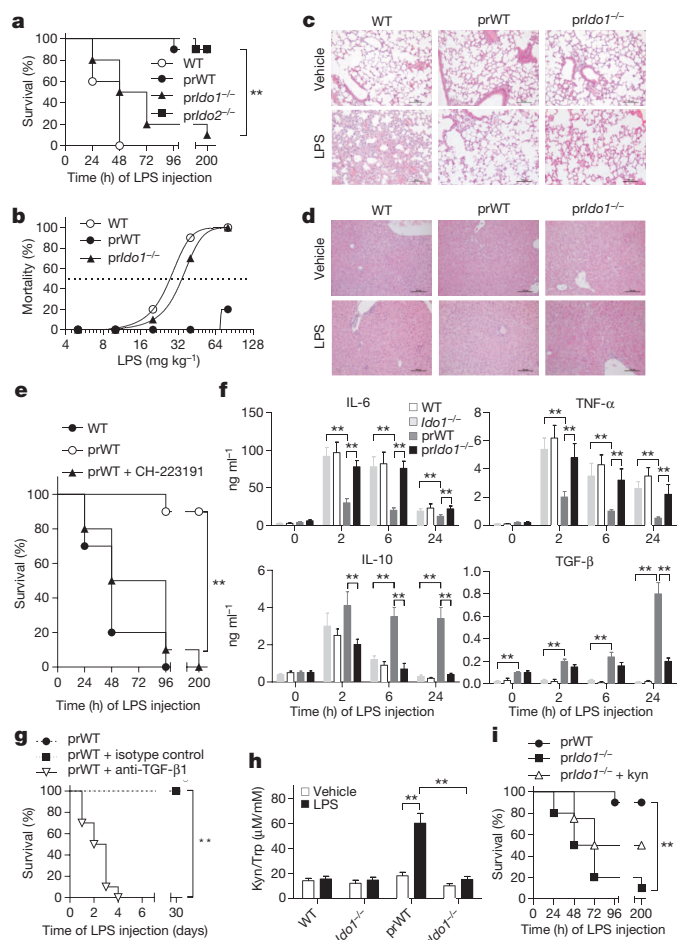


Figure 3 | Absolute requirement for AhR, functional IDO1 and TGF- β in LPS tolerance. **a**, Survival of WT and LPS-primed WT (prWT), IDO1-deficient (prIdo1^{-/-}) and IDO2-deficient (prIdo2^{-/-}) mice after rechallenge. $n = 8$ –10; one experiment of three. $^{**}P < 0.001$; log-rank test. **b**, Endotoxin LD₅₀ in tolerized mice. $n = 10$. **c, d**, Histopathology in lungs and liver, respectively; scale bar, 100 μ m. One experiment of two. **e**, Survival of variously treated mice ($n = 10$; one of three experiments). $^{**}P < 0.001$ (log-rank test). **f**, Cytokine measurements (mean \pm s.d. of three experiments; $n = 6$). $^{**}P < 0.001$ (two-tailed Student's t -test). **g**, Neutralization of TGF- β upon LPS rechallenge ($n = 10$). Three experiments (mean \pm s.d.); $^{**}P < 0.001$ (log-rank test). **h**, Kyn/Trp ratios (mean \pm s.d. of three experiments). $^{**}P < 0.001$ (two-tailed Student's t -test). **i**, Survival of variously treated mice. $n = 8$ –10; one experiment of three. $^{**}P < 0.001$ (log-rank test).

suggesting a necessary role for IDO1-dependent tryptophan catabolism in AhR-dependent gene transcription under those conditions of LPS rechallenge (Extended Data Fig. 4c). Likewise, the AhR requirement for endotoxin tolerance was demonstrated by the use of the specific antagonist, CH-223191 (Fig. 3e).

We also explored whether *Ahr*^{-/-} mice could be tolerized by sublethal LPS amounts. The highest LPS dosage resisted by naive *Ahr*^{-/-} mice (that is, 0.5 mg per kg) would still confer some degree of endotoxin tolerance on wild-type mice but not on *Ahr*-deficient mice (Extended Data Fig. 5a). In contrast, TDO2 was not required for endotoxin tolerance, because inhibition of TDO2 at the time of LPS rechallenge with its selective inhibitor had no effects, and *Tdo2*^{-/-} survivors of a primary LPS challenge would resist rechallenge to an extent comparable to primed wild-type controls (Extended Data Fig. 5b). Thus endotoxin tolerance, once established, did not require functional TDO2, but was dependent on the combined effects of IDO1 and AhR.

The early pattern of cytokine production by LPS-tolerant mice (at 2–24 h of rechallenge) was characterized by progressively decreasing

levels of circulating TNF- α and IL-6, but sustained production of IL-10 and the appearance of circulating TGF- β , contingent on IDO1 (Fig. 3f). The cytokine secretion pattern of endotoxin-tolerant wild-type mice was in marked contrast with that of *Ido1*^{-/-} mice, which resembled the cytokine secretion profile of unprimed wild-type controls (Fig. 3f). TGF- β was required for endotoxin tolerance because neutralization of TGF- β at the time of LPS rechallenge ablated protection (Fig. 3g). Increased serum tryptophan-to-kynurenine ratios were observed, on rechallenge, only in IDO1-competent mice (Fig. 3h).

In vivo, we investigated whether supplementary L-kynurenine, similar to what was observed in the primary challenge of TDO2 KO mice with LPS, would be sufficient alone to reinstall tolerance in otherwise susceptible LPS-primed *Ido1*^{-/-} mice. We treated mice with L-kynurenine at 6 h of LPS rechallenge (Fig. 3i). L-kynurenine failed to restore tolerance in the absence of IDO1, and there occurred a marked defect in TGF- β production in those mice, indicating that IDO1 was required for circulating TGF- β production. Additional IDO1 functions, beside kynurenine production, might thus be required for establishing a fully protective endotoxin-tolerant state.

AhR-dependent IDO1 phosphorylation

IDO1-dependent effects include non-enzymic functions, namely intracellular signalling events that, initiated by phosphorylation of specific domains in the enzyme, are involved in reprogramming gene expression and in the induction of a stably regulatory phenotype in splenic pDCs²⁶. Such polarized, tolerogenic pDCs produce less IL-6, but they start producing TGF- β in response to TLR signalling²⁷ or to TGF- β *in vitro*²⁸. Because of the failure of supplementary L-kynurenine to confer significant endotoxin tolerance on LPS-primed *Ido1*^{-/-} mice, we investigated whether IDO1-mediated signalling contributed to LPS tolerance, and in particular, whether IDO1 phosphorylation occurred in LPS-primed, IDO1-competent mice on rechallenge *in vivo* with LPS. We found that IDO1 phosphorylation could be detected under *ex vivo* conditions, in isolated spleen cells mostly expressing the CD11c marker, at 15 min of LPS rechallenge *in vivo* (Fig. 4a, b). Notably, IDO1 phosphorylation was not detected in the spleens of IDO1 KO mice. In accordance with the *ex vivo* data, AhR-dependent IDO1 phosphorylation could also be demonstrated in an *in vitro* model system using LPS-primed, IDO1-competent CD11c⁺ DCs on rechallenge with LPS (Fig. 4c).

To gain preliminary insight into the nature of the kinase responsible for the effect, we searched cDC gene expression profiles in public data sets, focusing on Src kinases known to mediate tyrosine-phosphorylation of intracellular target proteins. Profile analysis indicated that, among several possibilities including *Sykb*, *Blk*, *Frk*, *Zap70*, *Fgr*, *Fyn*, *Hck*, *Lck*, *Lyn*, *Src*, *Yes1* and *Csk*, the most widely represented kinase in mouse LPS-treated cDCs was the product of *Src* (Extended Data Fig. 6a, b). Src kinase activity is triggered by AhR activation^{29,30}, resulting in autophosphorylation and phosphorylation of target proteins³¹. Early, Src-dependent phosphorylation of IDO1 in primed cDCs occurred in response to LPS, and inhibition of Src kinase using a selective inhibitor, PP2 (but not its inactive analogue, PP3) negated IDO1 phosphorylation in LPS-primed cDCs re-exposed to LPS (Fig. 4c). Also, Src activity required AhR (Fig. 4d), which was necessary for TGF- β production (Fig. 4e). Overall, these data indicated that, in endotoxin tolerance, AhR-associated Src activity is responsible for IDO1 phosphorylation and TGF- β production by IDO1-competent cDCs.

LPS tolerance prevention of immunopathology

TLR4 and TLR2 activate innate immune cells in response to Gram-negative and Gram-positive bacteria, respectively. We analysed the impact of endotoxin tolerance on acute infection with LPS-expressing *Salmonella enterica* Typhimurium (S. Typhimurium, which exploits intestinal inflammation for pathogenicity³²; Fig. 5), as well as on group B *Streptococcus* (GBS)-induced multifocal septic arthritis, a condition in which functional TLR2 signalling is necessary for attenuating the severity

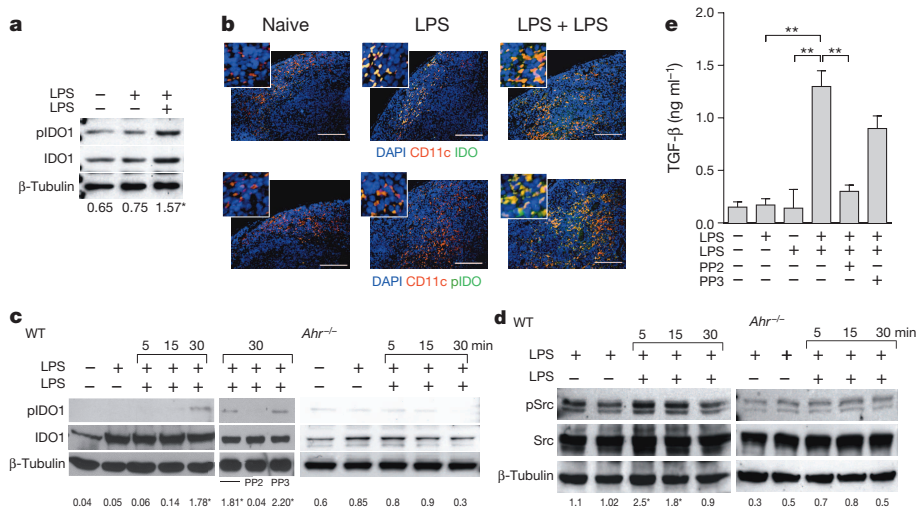


Figure 4 | LPS tolerance induces AhR- and Src-dependent IDO1 phosphorylation. **a**, Phosphorylation of IDO1 in total splenocytes isolated from LPS-tolerant mice in one experiment of three. pIDO1/IDO1 ratios shown below the gels are mean values from the three experiments. **P* < 0.001 (15 min versus single LPS exposure; two-tailed Student's *t*-test). **b**, Splenic immunofluorescent staining. Scale bar, 100 μ m. **c**, Phosphorylation of IDO1's ITIM2 in WT and AhR-deficient cDCs. One experiment of three, ratios being means from the three experiments. **P* < 0.001 (30 min versus single LPS exposure; two-tailed Student's *t*-test). **d**, Immunoblot analysis of Src and phosphorylated Src-Y416 (pSrc) in WT and AhR-deficient cDCs in one experiment of three. Ratios are means of the three experiments. **P* < 0.001 (5 min and 15 min versus single LPS exposure; two-tailed Student's *t*-test). **e**, Production of TGF- β by Src-competent cDCs in response to single or double LPS exposure. Means \pm s.d. from three experiments; ***P* < 0.001 (two-tailed Student's *t*-test).

of group B streptococcal disease³³ (Fig. 6). In the *Salmonella* model, both clinical and serological data (Fig. 5a–c), as well as histopathology (Fig. 5d, e), provided evidence for reduced inflammatory pathology in LPS-tolerant mice, which was confirmed by lower levels of IL-1 β and TNF- α in caecum cell supernatants (Extended Data Fig. 7a). Moreover, higher levels of IL-10 and TGF- β were found in those mice, in which *Foxp3* expression in mesenteric lymph nodes was enhanced, as opposed to *Rorc* expression (Extended Data Fig. 7a). A protective role for AhR and IDO1 could also be demonstrated in naive mice (Extended Data Fig. 8a–c). On assessing the role of endotoxin tolerance in group B *Streptococcus*-induced septic arthritis, LPS-tolerant mice were infected intravenously with type IV group B streptococci, and disease severity was monitored by assessing survival rates (Fig. 6a), clinical scores of arthritis (Fig. 6b), bacterial burden (Fig. 6c), histopathology (Fig. 6d), and proinflammatory cytokine production (Fig. 6e). An overall strongly protective effect of LPS tolerance was observed against *Streptococcus*-induced immunopathology (Extended Data Fig. 9), and again, a protective role for AhR and IDO1 could also be demonstrated in naive mice (Extended Data Fig. 10a–c).

Discussion

Several considerations may apply to our current findings. First, our data indicate that, although the class of an immune response is tailored to fit the invading pathogen, its regulation is primarily tailored to fit the tissue in which the response occurs^{34,35}. Second, endotoxin tolerance can be seen as a form of protective³⁶ or infectious³⁷ tolerance, whereby the host organism protects itself from infectious diseases by reducing the negative impact of infections on host fitness, an ability epitomized in ref. 38 as “disease tolerance”. Whereas disease resistance aims at reducing or eliminating infection, disease tolerance is confirmed here to imply an ability to reduce the impact of infection on host fitness. Third, the IDO1 mechanism seems to have been selected through phylogenesis primarily to prevent overreacting responses to TLR-recognized pathogen-associated molecular patterns³⁹, and only later did it become involved in the response to T-cell receptor-recognized antigen^{15,26,40–43}. AhR, in turn, which has a role in regulatory T-cell generation¹⁷, is presumed to have evolved from invertebrates, where it served a ligand-independent role in normal development processes. Evolution of the receptor in vertebrates resulted in the ability to bind structurally different ligands, including xenobiotics and microbiota-derived catabolites⁴⁴. Considering the inability of invertebrate AhR homologues to bind dioxins, the adaptive role of the AhR to act as a regulator of xenobiotic-metabolizing enzymes may have been a vertebrate innovation, to later

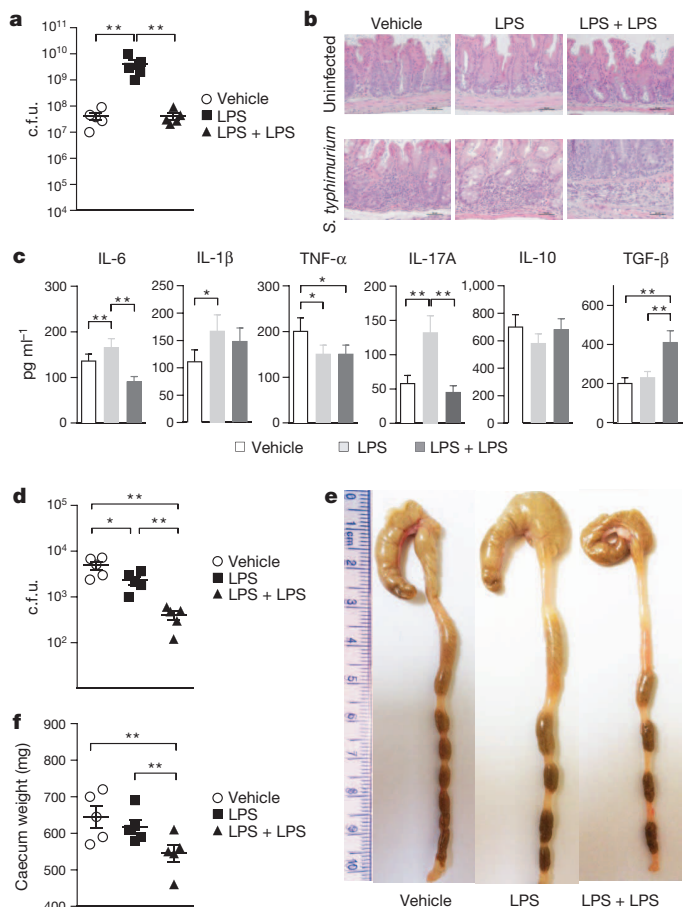


Figure 5 | LPS tolerance ameliorates *S. Typhimurium* infection. **a**, Mice, treated once or twice with LPS, were challenged with *S. Typhimurium*, and bacterial counts assessed in caecum (c.f.u., colony-forming units); ***P* < 0.001; two-tailed Student's *t*-test; **b**, Haematoxylin and eosin staining of caecal tissues; **c**, Cytokines in caecum cell supernatants. Data are means (\pm s.d.) from three experiments; **P* < 0.05, ***P* < 0.01; two-tailed Student's *t*-test. **d**, Caecal c.f.u. and caecal weights at 14 days post infection in tolerant mice. Three experiments; means \pm s.d., with **P* < 0.05 and ***P* < 0.001; two-tailed Student's *t*-test. **e**, Gross histopathology of caecum and distal colon (1-cm ruler scale).

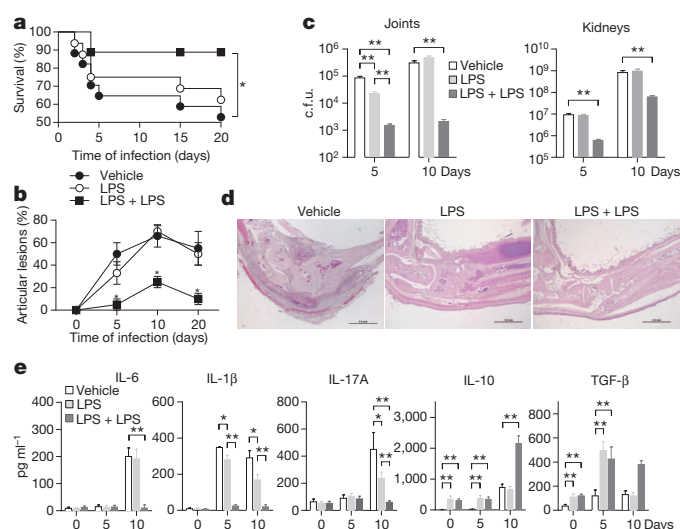


Figure 6 | LPS tolerance protects against *Streptococcus* immunopathology. **a**, Survival of variously treated, infected mice ($*P < 0.05$; log-rank test). **b**, Arthritis index. **c**, c.f.u. enumeration in joints and kidneys. In **b** and **c**, values are means \pm s.d. of three experiments; $*P < 0.05$ and $**P < 0.001$; two-tailed Student's *t*-test. **d**, Histopathologic evaluation of arthritis in joints from LPS-tolerant, GBS-infected mice. Representative data from one of three experiments. Scale bar, 1 mm. **e**, Supernatants from joint homogenates were assayed kinetically for IL-6, IL-1 β , IL-17A, IL-10 and TGF- β content. Values are the mean \pm s.d. of three experiments. $*P < 0.05$ and $**P < 0.001$, two-tailed Student's *t*-test.

acquire an additional immune regulatory role by coevolutionary pressure in mammals by IDO1 and regulatory T cells^{41–44}. Although *Ahr*^{−/−} mice have already been described as manifesting increased susceptibility to endotoxaemia¹⁰, our data on the respective inability of TDO2 KO and IDO1 KO mice to mount protective responses to LPS challenge and rechallenge may be the first to document uncompensated immune deficits in those mice. AhR would, indeed, engage in the long-term regulation of systemic, TLR-triggered inflammation through IDO1, resulting in the onset of a fully protective endotoxin-tolerant state, as exemplified by the *Salmonella* infection model. However, it is possible that the immunoregulatory effects of tryptophan catabolism represent a generalized phenomenon in infection and infection-related pathology^{45,46}. Our findings in the *Streptococcus*-induced septic arthritis model demonstrate that endotoxin tolerance does not prevent the early events in protective TLR2 signalling and that it may, in fact, accelerate the later onset of the host's control over immunologically mediated local and systemic disease. This would suggest that endotoxin tolerance helps to control specific, inflammasome-mediated effects of bacterial infection⁴⁷. Interfering with the trade-offs that the defence systems in both pathogens and their hosts impose on host fitness could ultimately help to selectively enhance pathways responsible for strong and rapid inflammatory responses that lead to effective pathogen clearance, yet do not involve immunopathology⁴⁸. Our current data may pave the way to new therapeutic options in infectious diseases via control of host-pathogen interactions. Druggable candidate targets in those interactions include sensors of pathogen presence (such as TLRs), alterations of host metabolism (including tryptophan catabolism) and environment-responsive receptors, of which AhR is a prototypical example.

METHODS SUMMARY

All *in vivo* studies were in compliance with National (Italian Parliament DL 116/92) and Perugia University Animal Care and Use Committee guidelines, and the overall study was approved by the Bioethics Committee of the University of Perugia. LPS LD₅₀ values for each mouse genotype were calculated by curve-fitting ($r^2 > 0.96$ – 0.99) and data are from one experiment representative of two or three. In tolerance experiments, mice were injected intraperitoneally with a first dose of LPS (10 mg per kg) and, after 6 days, rechallenged with 40 mg per kg LPS. Determination

of L-kynurenine K_i for AhR was done by a ligand binding assay by measuring the binding fixed concentration (0.3 nM) of [³H]TCDD at equilibrium in the presence of an incrementing series of concentrations of unlabelled L-kynurenine. The equilibrium inhibitor constant K_i was calculated based on the formula, $K_i = IC_{50}/(1 + [L]/K_{dL})$ where L designates the radioligand. Real-time PCR with reverse transcription (for mouse genes *Ido1*, *Ido2*, *Tdo2*, *Ahr*, *Il10*, *Tgfb1* and *Gapdh*) analyses were carried out as described previously²⁸, and data are presented as normalized transcript expression in the samples relative to normalized transcript expression in control cultures. For molecular modelling, a multiple alignment between the sequences of PAS-B mAhR (residues 278–384) and PAS-B HIF-2 α was generated according to the sequence alignment suggested in ref. 24. Phosphorylation of IDO1 *ex vivo* was examined in total splenocytes isolated from wild-type mice, either unprimed or primed *in vivo* with LPS (10 mg per kg), with or without rechallenge with 40 mg per kg LPS on day 7 of priming. For examining expression of tyrosine kinases in mouse myeloid dendritic cells, we retrieved seven data sets of mouse myeloid dendritic cells from Gene Expression Omnibus, containing information about treatment with LPS and gene expression data obtained with Affymetrix Mouse Genome 430 2.0 arrays.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 31 January; accepted 10 April 2014.

Published online 15 June; corrected online 9 July 2014 (see full-text HTML version for details).

- Fan, H. & Cook, J. A. Molecular mechanisms of endotoxin tolerance. *J. Endotoxin Res.* **10**, 71–84 (2004).
- Pena, O. M., Pistolic, J., Raj, D., Fjell, C. D. & Hancock, R. E. Endotoxin tolerance represents a distinctive state of alternative polarization (M2) in human mononuclear cells. *J. Immunol.* **186**, 7243–7254 (2011).
- Krausgruber, T. *et al.* IRF5 promotes inflammatory macrophage polarization and T_H1-T_H17 responses. *Nature Immunol.* **12**, 231–238 (2011).
- Abdi, K., Singh, N. J. & Matzinger, P. Lipopolysaccharide-activated dendritic cells: “exhausted” or alert and waiting? *J. Immunol.* **188**, 5981–5989 (2012).
- Biswas, S. K. & Lopez-Collazo, E. Endotoxin tolerance: new mechanisms, molecules and clinical significance. *Trends Immunol.* **30**, 475–487 (2009).
- Park, S. H., Park-Min, K. H., Chen, J., Hu, X. & Ivashkiv, L. B. Tumor necrosis factor induces GSK3 kinase-mediated cross-tolerance to endotoxin in macrophages. *Nature Immunol.* **12**, 607–615 (2011).
- Doreswamy, V. & Peden, D. B. Modulation of asthma by endotoxin. *Clin. Exp. Allergy* **41**, 9–19 (2011).
- Stejskalova, L., Dvorak, Z. & Pavlek, P. Endogenous and exogenous ligands of aryl hydrocarbon receptor: current state of art. *Curr. Drug Metab.* **12**, 198–212 (2011).
- Quintana, F. J. The aryl hydrocarbon receptor: a molecular pathway for the environmental control of the immune response. *Immunology* **138**, 183–189 (2013).
- Kimura, A. *et al.* Aryl hydrocarbon receptor in combination with Stat1 regulates LPS-induced inflammatory responses. *J. Exp. Med.* **206**, 2027–2035 (2009).
- Nguyen, L. P. & Bradfield, C. A. The search for endogenous activators of the aryl hydrocarbon receptor. *Chem. Res. Toxicol.* **21**, 102–116 (2008).
- Murray, M. F. The human indoleamine 2,3-dioxygenase gene and related human genes. *Curr. Drug Metab.* **8**, 197–200 (2007).
- Orabona, C. *et al.* Toward the identification of a tolerogenic signature in IDO-competent dendritic cells. *Blood* **107**, 2846–2854 (2006).
- Stone, T. W., Stoy, N. & Darlington, L. G. An expanding range of targets for kynurenine metabolites of tryptophan. *Trends Pharmacol. Sci.* **34**, 136–143 (2013).
- Fallarino, F. *et al.* The combined effects of tryptophan starvation and tryptophan catabolites down-regulate T cell receptor ζ -chain and induce a regulatory phenotype in naive T cells. *J. Immunol.* **176**, 6752–6761 (2006).
- Nguyen, N. T. *et al.* Aryl hydrocarbon receptor negatively regulates dendritic cell immunogenicity via a kynurenine-dependent mechanism. *Proc. Natl Acad. Sci. USA* **107**, 19961–19966 (2010).
- Mezrich, J. D. *et al.* An interaction between kynurenine and the aryl hydrocarbon receptor can generate regulatory T cells. *J. Immunol.* **185**, 3190–3198 (2010).
- Romani, L. *et al.* Defective tryptophan catabolism underlies inflammation in mouse chronic granulomatous disease. *Nature* **451**, 211–215 (2008).
- Changsirivathanathamrong, D. *et al.* Tryptophan metabolism to kynurenine is a potential novel contributor to hypotension in human sepsis. *Crit. Care Med.* **39**, 2678–2683 (2011).
- Jung, I. D. *et al.* Blockade of indoleamine 2,3-dioxygenase protects mice against lipopolysaccharide-induced endotoxin shock. *J. Immunol.* **182**, 3146–3154 (2009).
- Sekine, H. *et al.* Hypersensitivity of aryl hydrocarbon receptor-deficient mice to lipopolysaccharide-induced septic shock. *Mol. Cell. Biol.* **29**, 6391–6400 (2009).
- Trifari, S., Kaplan, C. D., Tran, E. H., Crellin, N. K. & Spits, H. Identification of a human helper T cell population that has abundant production of interleukin 22 and is distinct from T_H17, T_H1 and T_H2 cells. *Nature Immunol.* **10**, 864–871 (2009).
- Howard, G. J., Schlezinger, J. J., Hahn, M. E. & Webster, T. F. Generalized concentration addition predicts joint effects of aryl hydrocarbon receptor agonists with partial agonists and competitive antagonists. *Environ. Health Perspect.* **118**, 666–672 (2009).

24. Pandini, A. *et al.* Detection of the TCDD binding-fingerprint within the Ah receptor ligand binding domain by structurally driven mutagenesis and functional analysis. *Biochemistry* **48**, 5972–5983 (2009).
25. Opitz, C. A. *et al.* An endogenous tumour-promoting ligand of the human aryl hydrocarbon receptor. *Nature* **478**, 197–203 (2011).
26. Fallarino, F., Grohmann, U. & Puccetti, P. Indoleamine 2,3-dioxygenase: from catalyst to signaling function. *Eur. J. Immunol.* **42**, 1932–1937 (2012).
27. De Luca, A. *et al.* Functional yet balanced reactivity to *Candida albicans* requires TRIF, MyD88, and IDO-dependent inhibition of *Rorc*. *J. Immunol.* **179**, 5999–6008 (2007).
28. Pallotta, M. T. *et al.* Indoleamine 2,3-dioxygenase is a signaling protein in long-term tolerance by dendritic cells. *Nature Immunol.* **12**, 870–878 (2011).
29. Dong, B. *et al.* FRET analysis of protein tyrosine kinase c-Src activation mediated via aryl hydrocarbon receptor. *Biochim. Biophys. Acta* **1810**, 427–431 (2011).
30. Randi, A. S. *et al.* Hexachlorobenzene triggers AhR translocation to the nucleus, c-Src activation and EGFR transactivation in rat liver. *Toxicol. Lett.* **177**, 116–122 (2008).
31. Backlund, M. & Ingelman-Sundberg, M. Regulation of aryl hydrocarbon receptor signal transduction by protein tyrosine kinases. *Cell. Signal.* **17**, 39–48 (2005).
32. Thiennimitr, P. *et al.* Intestinal inflammation allows *Salmonella* to use ethanolamine to compete with the microbiota. *Proc. Natl Acad. Sci. USA* **108**, 17480–17485 (2011).
33. Puliti, M., Uematsu, S., Akira, S., Bistoni, F. & Tissi, L. Toll-like receptor 2 deficiency is associated with enhanced severity of group B streptococcal disease. *Infect. Immun.* **77**, 1524–1531 (2009).
34. Matzinger, P. & Kamala, T. Tissue-based class control: the other side of tolerance. *Nature Rev. Immunol.* **11**, 221–230 (2011).
35. Sander, L. E. *et al.* Hepatic acute-phase proteins control innate immune responses during infection by promoting myeloid-derived suppressor cell function. *J. Exp. Med.* **207**, 1453–1464 (2010).
36. Romani, L. & Puccetti, P. Protective tolerance to fungi: the role of IL-10 and tryptophan catabolism. *Trends Microbiol.* **14**, 183–189 (2006).
37. Belladonna, M. L., Orabona, C., Grohmann, U. & Puccetti, P. TGF- β and kynurenines as the key to infectious tolerance. *Trends Mol. Med.* **15**, 41–49 (2009).
38. Medzhitov, R., Schneider, D. S. & Soares, M. P. Disease tolerance as a defense strategy. *Science* **335**, 936–941 (2012).
39. Volpi, C. *et al.* High doses of CpG oligodeoxynucleotides stimulate a tolerogenic TLR9-TRIF pathway. *Nature Commun.* **4**, 1852 (2013).
40. Grohmann, U. *et al.* CTLA-4-Ig regulates tryptophan catabolism *in vivo*. *Nature Immunol.* **3**, 1097–1101 (2002).
41. Fallarino, F. *et al.* Modulation of tryptophan catabolism by regulatory T cells. *Nature Immunol.* **4**, 1206–1212 (2003).
42. Munn, D. H. *et al.* Prevention of allogeneic fetal rejection by tryptophan catabolism. *Science* **281**, 1191–1193 (1998).
43. Samstein, R. M., Josefowicz, S. Z., Arvey, A., Treuting, P. M. & Rudensky, A. Y. Extrathymic generation of regulatory T cells in placental mammals mitigates maternal-fetal conflict. *Cell* **150**, 29–38 (2012).
44. Zelante, T. *et al.* Tryptophan catabolites from microbiota engage aryl hydrocarbon receptor and balance mucosal reactivity via interleukin-22. *Immunity* **39**, 372–385 (2013).
45. Zelante, T., Fallarino, F., Bistoni, F., Puccetti, P. & Romani, L. Indoleamine 2,3-dioxygenase in infection: the paradox of an evasive strategy that benefits the host. *Microbes Infect.* **11**, 133–141 (2009).
46. Chen, W. IDO: more than an enzyme. *Nature Immunol.* **12**, 809–811 (2011).
47. Martinon, F., Mayor, A. & Tschopp, J. The inflammasomes: guardians of the body. *Annu. Rev. Immunol.* **27**, 229–265 (2009).
48. Chambers, M. C. & Schneider, D. S. Balancing resistance and infection tolerance through metabolic means. *Proc. Natl Acad. Sci. USA* **109**, 13886–13887 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by funding from the Italian Association for Cancer Research (AIRC, to P.P.), Fondazione Italiana Sclerosi Multipla Project No. 2010/R/17 (to F.F.), Associazione Umbra Contro il Cancro (to G.S. & M.A.D.F.), Bayer Grants4Target Focus Grant no. 2012-03-0630 (to A.I., F.F. and D.M.), Bayer Early Career Investigator Award (to D.M.), Grant no. R01ES007685 from the US National Institutes of Environmental Health Sciences (to M.S.D.), the Specific Targeted Research Project FUNMETA (to L.R.), and the Italian Ministry of Health in association with Regione dell'Umbria (GR-2008-1138004 to C.O.) We thank D. Fuchs for serum kynurenine determinations. We also thank G. Andrielli for digital art and image editing and G. Ricci for histopathology.

Author Contributions A.B. and M.G. designed and conducted all experiments unless otherwise indicated below; M.T.P., D.M. and C.V. analysed IDO and Src phosphorylation; S.B., E.M.C.M., D.P., M.P. and A.I. conducted bioinformatics studies and statistical analysis. A.M. performed homology modelling and docking studies. R.I., T.Z., M.A.D.F., L.R. and L.T. conducted the *in vivo* studies with *Salmonella* and GBS. C.V., M.L.B., C.O., G.S., C.B. and R.B. contributed to specific experimental designs; T.V.L., M.P., H.F. and T.N. made possible, and designed, the experiments with TDO2-deficient mice; J.B.D., G.C.P. and R.M. made possible, and designed, the experiments with IDO2-deficient mice; D.G., M.S.D., G.J.G., M.G., B.V., L.B. and U.G. provided conceptual help and reagents throughout experimentation; F.F. designed and supervised all experiments; P.P. supervised the overall study and wrote the manuscript. F.F. and P.P. share senior authorship on this paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.P. (plopcc@tin.it) or F.F. (flfnc@tin.it).

NMDA receptor structures reveal subunit arrangement and pore architecture

Chia-Hsueh Lee^{1*}, Wei Lü^{1*}, Jennifer Carlisle Michel^{1,2}, April Goehring^{1,2}, Juan Du¹, Xianqiang Song¹ & Eric Gouaux^{1,2}

N-methyl-D-aspartate (NMDA) receptors are Hebbian-like coincidence detectors, requiring binding of glycine and glutamate in combination with the relief of voltage-dependent magnesium block to open an ion conductive pore across the membrane bilayer. Despite the importance of the NMDA receptor in the development and function of the brain, a molecular structure of an intact receptor has remained elusive. Here we present X-ray crystal structures of the *Xenopus laevis* GluN1–GluN2B NMDA receptor with the allosteric inhibitor, Ro25-6981, partial agonists and the ion channel blocker, MK-801. Receptor subunits are arranged in a 1–2–1–2 fashion, demonstrating extensive interactions between the amino-terminal and ligand-binding domains. The transmembrane domains harbour a closed-blocked ion channel, a pyramidal central vestibule lined by residues implicated in binding ion channel blockers and magnesium, and a ~twofold symmetric arrangement of ion channel pore loops. These structures provide new insights into the architecture, allosteric coupling and ion channel function of NMDA receptors.

Glutamate is the primary excitatory neurotransmitter in the brain, acting at ionotropic and metabotropic glutamate receptors. Rapid excitation by glutamate, in turn, solely involves action at AMPA, kainate and NMDA ionotropic glutamate receptors¹. The NMDA receptor is central to the development and function of the nervous system and to neurotoxicity². As a linchpin of synaptic plasticity, blockade of the NMDA receptor interferes with memory formation and recall³. Moreover, mutations within the coding regions of NMDA receptor subunit genes are associated with a spectrum of neurological diseases and neuropsychiatric disorders, from schizophrenia to epilepsy⁴. Autoimmune responses to the NMDA receptor, and presumed disruption in NMDA receptor organization on neural cell surfaces, probably underlie NMDA receptor encephalitis⁵. In keeping with the profound roles of the NMDA receptor in brain function, the receptor is a target of small molecules for the treatment of cognitive impairment, depression, schizophrenia and pain².

Although AMPA and kainate receptors can be activated solely by glutamate^{6–8}, NMDA receptors are Hebbian-like coincidence detectors, requiring the binding of glycine and glutamate to GluN1 and GluN2 subunits⁹, respectively, combined with membrane depolarization to relieve magnesium block^{10,11}. Activation of the receptor opens a cation-selective, calcium-permeable channel, thus causing further depolarization of the cell membrane and influx of calcium¹². NMDA receptors are obligatory heterotetrameric assemblies^{13,14}, usually composed of two glycine-binding GluN1 subunits and two glutamate-binding GluN2A–D subunits, with the GluN1–GluN2A–GluN2B complex as the predominant NMDA receptor at hippocampal synapses¹⁵. Glycine- and D-serine-binding GluN3 subunits are additional subunits, expressed throughout the nervous system, but with roles less well defined in comparison to the GluN1–GluN2 assemblies. A hallmark of NMDA receptors, by contrast with AMPA and kainate receptors, is a wide spectrum of allosteric modulation, from nanomolar concentrations of zinc, to the small molecule ifenprodil, polyamines and protons¹⁶ and to voltage-dependent ion channel block by MK-801, ketamine and memantine¹⁷.

The GluN1, GluN2 and GluN3 NMDA receptor subunits are related in amino acid sequence and, like AMPA and kainate receptor subunits,

possess a modular domain architecture, with amino-terminal domains (ATDs) and ligand-binding domains (LBDs) on the extracellular side of the membrane, a transmembrane domain (TMD) spanning the membrane and defining the ion channel pore, and an intracellular carboxy-terminal domain (CTD) within the cytoplasm¹. Multiple high-resolution crystal structures of the isolated LBDs from NMDA, AMPA and kainate receptors show that these domains adopt similar clamshell-like structures that are organized in an approximately dimeric, back-to-back fashion^{18–20}. Although crystal structures of isolated ATDs illustrate that they too possess a clamshell-like structure^{21–23}, in NMDA receptors not only is the organization of each clamshell lobe distinct from that in AMPA and kainate receptors, but the interactions between subunits are also different²⁴. The functional properties of the NMDA ion channel pore, which harbours binding sites for magnesium and small molecule blockers, are also distinct from AMPA and kainate receptors¹.

Here we report crystal structures of the GluN1–GluN2B NMDA receptor from *Xenopus laevis* in complex with the GluN2B-specific allosteric inhibitor, Ro25-6981 (ref. 25), the GluN1 and GluN2B partial agonists 1-aminocyclopropane-1-carboxylic acid (ACPC)²⁶ and *trans*-1-aminocyclobutane-1,3-dicarboxylic acid (t-ACBD)²⁷, respectively, and the ion channel blocker, MK-801. To enhance the stability of the receptor in detergent micelles and to reduce conformational surface entropy, we replaced the cytoplasmic C terminus of the GluN1 and GluN2B subunits with 11 residues from the GluA2 C terminus²⁸ and we introduced a number of mutations into each subunit, ultimately finding a NMDA receptor complex that preserved binding of full and partial agonists and Ro25-6981, together with small but measurable conductance activated by glycine and glutamate, and with channel block by magnesium. To decrease conformational mobility of the extracellular domains, we substituted GluN2B Lys 216 to Cys (K216C), resulting in spontaneous disulphide bond formation between GluN2B subunits, improving crystal quality yet reducing agonist-induced ion channel activity (Extended Data Table 1 and Extended Data Figs 1–4). We determined crystal structures of the GluN1–GluN2B K216C receptor at resolutions of 3.7 Å (structure 1) and 3.9 Å (structure 2) and refined the

¹Vollum Institute, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97239, USA. ²Howard Hughes Medical Institute, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97239, USA.

*These authors contributed equally to this work.

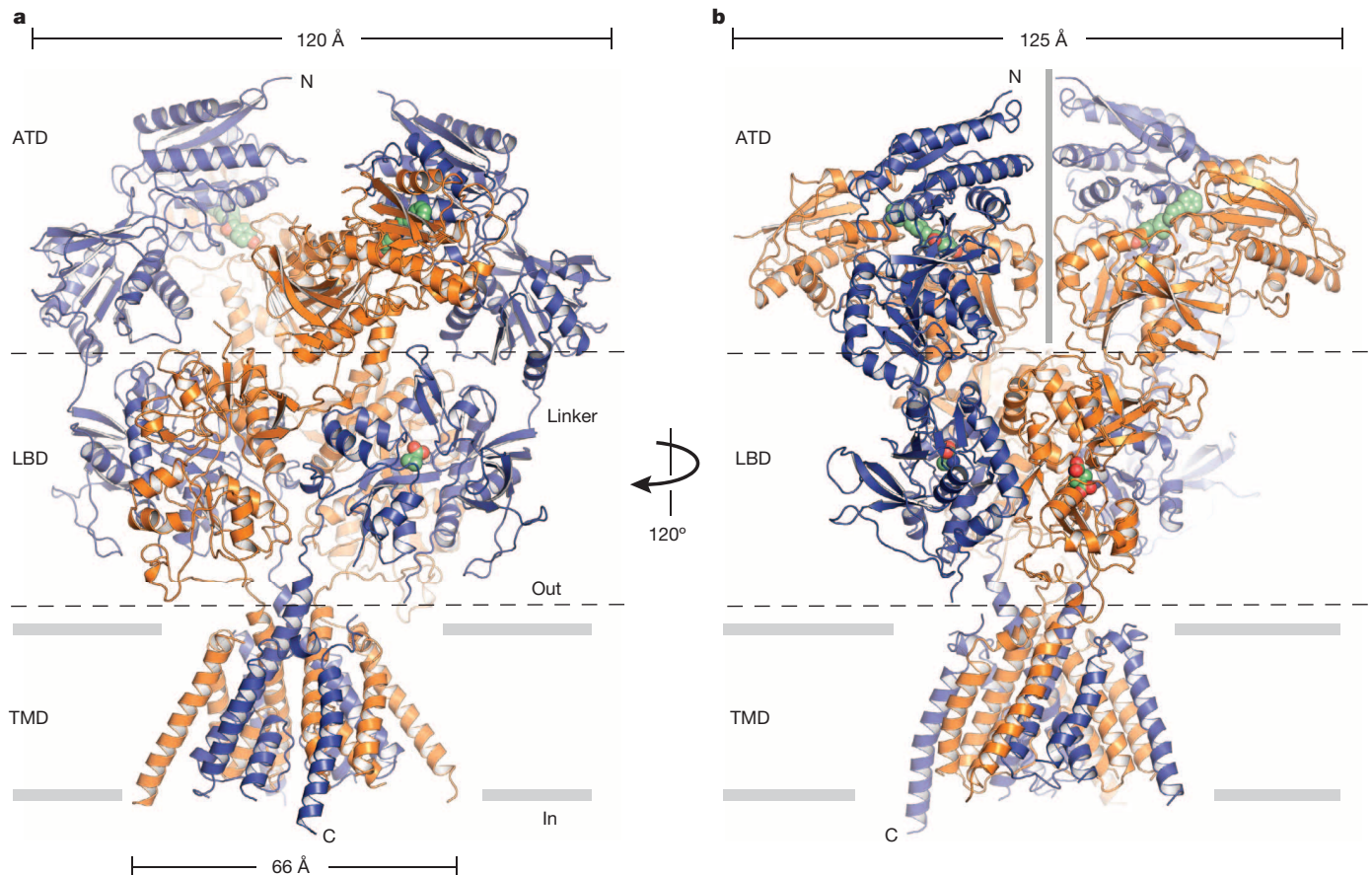


Figure 1 | Architecture, symmetry and domain organization of the GluN1–GluN2B NMDA receptor. **a**, View of the receptor complex, parallel to the membrane, with the GluN1 subunits in blue and the GluN2B subunits in orange. The ligands Ro25-6981, ACPC and t-ACBD are in space-filling

structures to reasonable crystallographic residuals and good stereochemistry. In addition, we mapped cation sites in the ATD by exploiting anomalous scattering from a Tb^{3+} derivative and probed the mobility of the ATD and LBD layers by comparing a non K216C crosslinked structure to the higher resolution K216C structures (Extended Data Table 2 and Supplementary Discussion).

Architecture and symmetry

The structure of the GluN1–GluN2B NMDA receptor resembles a mushroom, with a height of ~ 150 Å and widths of $\sim 125 \times 120$ Å (Fig. 1a, b and Supplementary Video 1). With an overall twofold axis of symmetry, reminiscent of the intact GluA2 AMPA receptor architecture and symmetry²⁸, the receptor domains are organized into three layers with the ATD layer at the top, the LBD layer in the middle and the TMD layer at the bottom. By contrast with the AMPA receptor, the extracellular layers are more compact, with the ATD layer adopting an entirely different structure, interdigitated within the crevices of the LBD layer. The LBD layer caps the extracellular end of the transmembrane domain, with loops from the GluN2B LBDs drooping towards the extracellular leaflet of the membrane bilayer (Fig. 1a). The TMD hews to an AMPA-like topology and arrangement of helices²⁸, yet with electron density for the M2 segments and pore loops in structure 2, allowing us to define the structure of nearly the entire ion channel pore. Structure 1 and structure 2 are similar, nevertheless, with an overall root mean square deviation (r.m.s.d.) on main chain atoms of 0.6 Å. Here we primarily use structure 1 to discuss the ATDs, LBDs and LBD to TMD linkages and structure 2 to describe the TMD.

Subunit arrangement within the GluN1–GluN2B NMDA receptor adheres to the organization of the AMPA receptor²⁸, with the glycine-binding GluN1 subunits occupying the A–C subunit positions and the

representation. **b**, View of the complex rotated $\sim 120^\circ$ around the overall twofold axis of the receptor. The approximate position of the overall twofold axis is shown by a vertical grey bar in the centre of the ATD layer. Structure 2 is shown.

glutamate-binding GluN2B subunits situated in the B–D subunit sites (Fig. 2a). In agreement with crosslinking studies on the GluN1–GluN2A receptor^{28–30} and isolated ATDs³¹ and in agreement with crystal structures of the GluN1–GluN2A LBDs³² and the GluN1–GluN2B ATDs²⁴, the ATDs and the LBDs are organized as local GluN1–GluN2B heterodimers. Like the AMPA receptor²⁸, there is subunit ‘crossover’ between ATD and LBD layers such that the subunits of a given ATD heterodimer are connected to subunits in a different LBD heterodimer, thus knitting together the extracellular domain superstructure. The TMDs are further stitched together by the M4 helices interacting nearly exclusively with TM segments from an adjacent subunit. The arrangement of subunits within this NMDA receptor complex illustrates how the subunit non-equivalence first described for the homomeric AMPA receptor²⁸ has been exploited in an obligatory heteromeric assembly.

Arrangement of amino-terminal domains

The GluN1–GluN2B ATDs are perched above the LBD layer, with the R2 lobes of the GluN2B subunits proximal to each other and near the overall twofold axis of symmetry while the GluN1 ATDs reside at the periphery of the receptor assembly (Fig. 2a, b). The ATD heterodimer is shaped like an inverted ‘V’, in which the open end of the ‘V’ straddles the GluN1 LBD beneath it, with loops and residues of the GluN1 R2 lobe interacting with its GluN1 LBD and the GluN2B R2 lobe wedged into the interdimer LBD interface. The conformation of an individual ATD heterodimer from the intact receptor structure is slightly contracted compared to the structure of the isolated heterodimer, perhaps because of interactions with the LBD layer or due to lattice contacts (Extended Data Fig. 5a, b and Extended Data Table 3). There is prominent electron density at the interface between the GluN1 and GluN2B subunits for

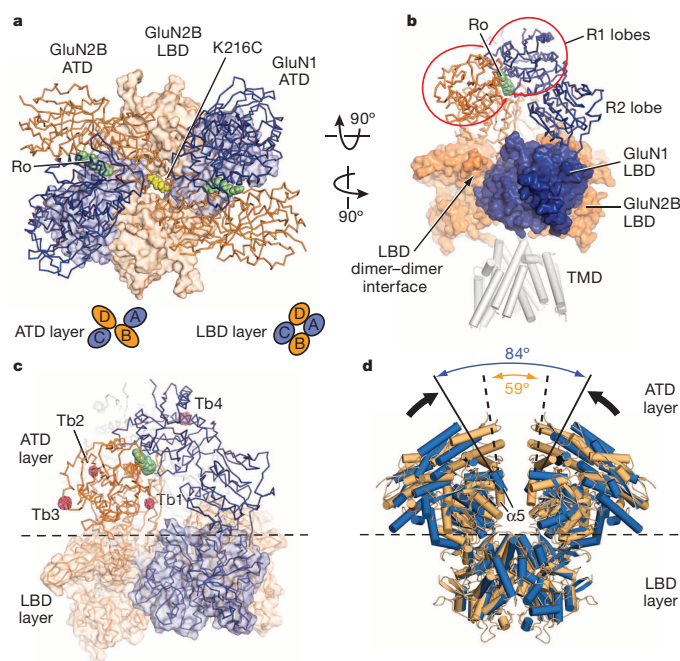


Figure 2 | ATD arrangement, cation binding sites and conformational mobility. **a**, View of the ATD layer along the overall twofold axis, from the extracellular side of the membrane, centred on the overall twofold axis, and showing the relative location of the underlying LBD layer. Ro25–6981 (Ro) is green and the K216C disulphide is yellow. The arrangements of subunits for ATD and LBD layers are shown as insets. **b**, The inverted ATD heterodimeric ‘V’ straddles GluN1 and GluN2B LBD subunits on different local LBD heterodimers. The ATD R2 lobes interact with the LBDs and the R1 lobes cradle bound Ro25–6981 at an ATD subunit interface. Structure 1 is shown in panels **a** and **b**. **c**, Tb³⁺ binding sites. An anomalous difference electron density map is shown contoured at 3 σ (pink mesh). Sites Tb1 and Tb2 are located at the ‘hinge’ between the R1 and R2 lobes, whereas sites Tb3 and Tb4 are at receptor–receptor contacts in the crystal lattice. **d**, The ATD and LBD extracellular domains derived from the two low-resolution GluN1–GluN2B receptor structures (Extended Data Table 2; data set 4/structure 4) where the GluN2B subunits do not harbour the K216C disulphide bridge, illustrating the conformational mobility of the ATD layer. The angles between the $\alpha 5$ helices of the GluN2B subunits for each of the two independent receptor complexes in the asymmetric unit illustrate the conformational mobility of the ATD layers.

the allosteric antagonist, Ro25–6981, where it stabilizes the intersubunit interface (Extended Data Fig. 5c)²⁴. A small interface, formed between the ATD heterodimers, is centred at the engineered disulphide crosslink at residue 216 of the $\alpha 5$ helix on the GluN2B R2 lobe (Fig. 2a).

Zinc acts as an antagonist at nanomolar concentrations on GluN2A-containing receptors and at micromolar concentrations on receptors harbouring the GluN2B subunit³³. Because lanthanum also antagonizes the NMDA receptor in a voltage-independent manner³⁴ and lanthanides can bind to zinc sites³⁵, we exploited the anomalous scattering signal of terbium and measured X-ray diffraction data near its f'' maximum. Anomalous difference electron density maps show two peaks near the interdomain hinge of the R1–R2 lobes of the GluN2B subunit (Fig. 2c and Extended Data Fig. 5d). The Tb2 site overlaps with the previously determined Zn1 site²³, whereas the other site (Tb1), near residues Glu 146, Asp 176 and Asp 349, is unique. These data support the notion that ions and small molecules can bind to the ATD clamshells³⁶ in a position to modulate ATD conformation, although future experiments are required to establish the roles of these sites in allosteric regulation of GluN2B-containing NMDA receptors.

The first structure of the NMDA receptor was derived from the low resolution data set 4 (Extended Data Table 2) and involved a construct lacking the GluN2B(K216C) mutant. In this crystal form, there are two halves of a receptor in the asymmetric unit and application of crystal

symmetry creates two intact receptors, each with a different conformation of the ATDs in which the angles of the ATD domains range from 59° to 84° across the overall twofold axis (Fig. 2d). We further observed that helix $\alpha 5$ of the GluN2B R2 lobes face each other, proximal to the overall twofold axis of symmetry. Because we speculated that these structures were indicative of substantial mobility in the ATD layer, we made single cysteine substitutions on the exposed face of helix $\alpha 5$ and screened for redox dependent crosslinking of GluN2B subunits. Indeed, the K216C mutant, as well as other residues on the face of $\alpha 5$, spontaneously form subunit–subunit crosslinks (Extended Data Figs 1, 2 and 4), bringing the GluN2B ATDs in close apposition (Extended Data Fig. 5e), diminishing ion channel activity and increasing the resolution to which the crystals diffract. In two electrode voltage clamp experiments, reduction of oocytes using dithiothreitol enhances current responses from the K216C mutant, suggesting that movements of the ATDs allosterically modulate the activity of the ion channel (Extended Data Fig. 2).

Ligand-binding-domain layer

The agonist-binding LBDs of the NMDA receptor are organized as a nearly equivalent pair of GluN1–GluN2B heterodimers in which each GluN1–GluN2B heterodimer (Fig. 3a, c) closely resembles the water-soluble heterodimers of the isolated GluN1–GluN2A LBDs³² and the homodimeric assemblies of AMPA¹⁸ and kainate receptor³⁷ LBDs in non-desensitized conformations. Moreover, the arrangement is similar to that previously observed in the structure of the full-length AMPA receptor (Fig. 3b, d)²⁸, although here the electron density for the GluN1 and GluN2B LBDs in chains B and C is weak, perhaps due to an absence of lattice contacts. In comparing this NMDA receptor structure to the antagonist-bound state of the AMPA receptor, the extent to which the local twofold axes of each LBD dimer are angled off of the overall molecular twofold axis of symmetry differ (Fig. 3a, b). In addition, inspection of the GluN1–GluN2B and AMPA receptor LBD layers, viewed from the top (Fig. 3c, d), shows that there is a relative translation, or shift, of the LBD dimers along the interdimer interface (Fig. 3e). Using helix J to align the B–C LBDs, the A–D LBD dimer in the AMPA receptor has undergone a translational ‘shift’ of ~15 Å relative to the A–D NMDA receptor LBD heterodimer. Although we do not know if these differences in LBD dimer ‘roll’ angle (Fig. 3a, b) and translational ‘shift’ (Fig. 3e) are due to inherent differences between NMDA and AMPA receptors or to the closed-blocked state of the NMDA receptor versus the competitive antagonist-bound form of the AMPA receptor, or to both factors, this analysis illustrates conformational mobility of the LBD dimers perhaps related to how the LBD couples agonist-binding to the TMD.

Within the LBD layer there are two major interfaces, one within a heterodimer, at the D1–D1 interface of GluN1 and GluN2B subunits, and the second between heterodimers. The intradimer D1–D1 interface is a region of allosteric modulation in NMDA receptors^{38,39} and, within one heterodimer, buries ~1,100 Å² of solvent accessible surface area on each subunit (Fig. 3c, f, boxed region). There are two nearly equivalent interdimer interfaces between the heterodimeric LBDs, each burying ~600 Å² of solvent-accessible surface area, and comprising two loci (Fig. 3a, boxed regions). One involves helix G on GluN1 interacting with residues on loop 1 of GluN2B (Fig. 3g) and the second includes residues on helix K (GluN2B) making contacts with residues on helix E (GluN1, Fig. 3h). Both sets of interactions involve contacts between residues on the D1 and D2 lobes of the GluN1 and GluN2B subunits, providing a direct route by which modulation of LBD clamshell closure could be translated into rearrangement of the LBD layer. As previously suggested, both the NMDA receptor LBD intradimer interface and the dimer–dimer interface may adopt different conformations depending on the functional state of the receptor.

The initial trigger for the eventual opening of the ion channel gate resides in agonist binding to the LBD clamshells. NMDA receptors require binding by agonists at both the GluN1 and GluN2 sites⁹, and here we have crystallized the receptor in complex with the partial agonists ACPC⁴⁰ and t-ACBD⁴¹. Agonist binding results in closure of the LBD

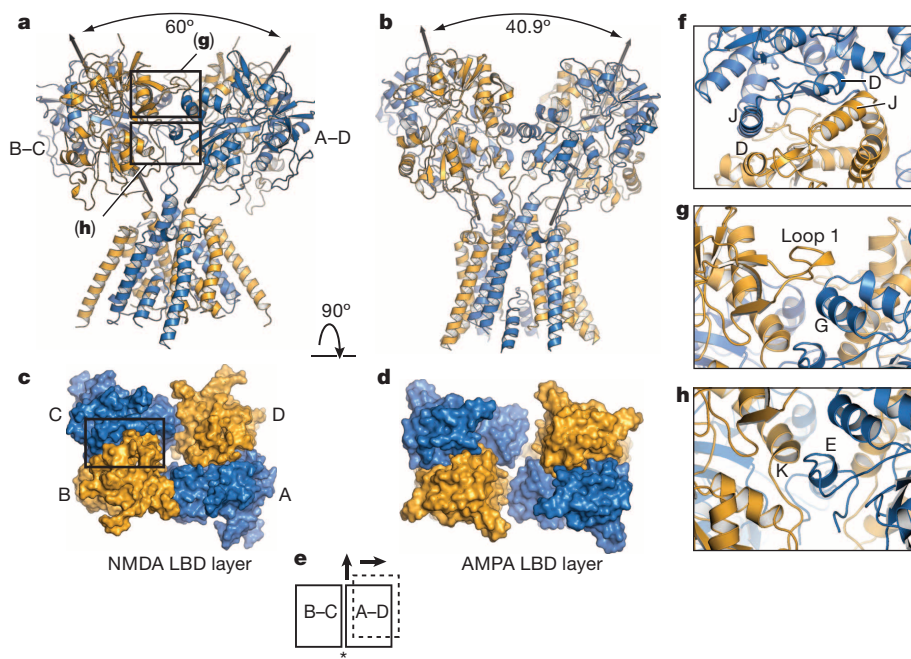


Figure 3 | LBD layer forms a ring-like structure. **a**, The GluN1–GluN2B LBD and TMD, showing that the pseudo twofold axes of the B–C and A–D LBD heterodimers diverge with an angle of 60° . The boxed areas define regions of LBD dimer–dimer contacts shown in panels **g** and **h**. **b**, View of the antagonist-bound state of the GluA2 AMPA receptor, which shows that the twofold axes of the LBD dimers diverge by an angle of 40.9° . **c**, View from the extracellular side of the membrane, along the overall twofold axis of the receptor, showing the LBDs of the GluN1 and GluN2B subunits, with the LBD heterodimer interface of the B–C subunits emphasized by a box. **d**, GluA2 LBD layer, illustrating how the interface between the B–C and A–D subunits has increased in comparison to the NMDA receptor LBD layer. **e**, Schematic of the LBD layer, showing the NMDA receptor B–C and A–D heterodimers as rectangles (solid lines) and illustrating the translational shift of the A–D subunits in the AMPA receptor (dotted lines). The asterisk indicates the dimer–dimer interface. **f–h**, Close-up view of the canonical D1–D1 intradimer interface³², together with views of the interactions at the inter-dimer interfaces in panels **g** and **h**. The domains from structure 2 are shown, with GluN1 subunits in blue and GluN2B subunits in orange.

clamshell⁴² and separation of the region proximal to the M3 transmembrane helix¹⁸. Analysis of the GluN1 and GluN2B LBDs demonstrates that each of the two GluN1 and GluN2B clamshells adopt similar conformations (Extended Data Table 3). Moreover, the degree of closure is similar to that observed for the isolated LBDs (Extended Data Fig. 6), except that they are both slightly more open in comparison to the isolated domains, perhaps owing to direct linkage to the ion channel. Separation of the region proximal to the M3 helices is similar between the equivalent residues in the LBD dimers of the full length receptor and in the glycine–glutamate complex of the isolated GluN1–GluN2A LBDs, yet longer than in an LBD antagonist 5,7-dichlorokynurenic acid (DCKA)–glutamate complex (Extended Data Fig. 6f–h). Thus, by this metric the LBD dimers adopt an agonist-bound, activated conformation.

ATD–LBD interactions and allosteric coupling

The molecular puzzle of how allosteric inhibitors such as Ro25-6981 and ifenprodil promote closure of the ion channel gate despite the binding of agonists to their cognate LBD dimers must be resolved, at least in part, through communication between the ATD and LBD layers and perturbation of the LBD layer from an active conformation to an inactive state. In the GluN1–GluN2B structure, the ATD heterodimers ‘straddle’ LBD subunits (Fig. 4a, b), with the R2 lobe of each GluN1 subunit making extensive contacts with the D1 lobe of the cognate GluN1 LBD and the R2 lobe of the GluN2B subunit insinuated into the LBD dimer–dimer interface, positioned to make contacts with its cognate GluN2B LBD and with the neighbouring GluN1 LBD. Thus we see that the ATDs are judiciously positioned to mediate conformational changes at both LBD intradimer^{18,38,39} and inter-dimer interfaces. Nevertheless, because the intradimer LBD interface is intact, the action of Ro25-6981 and related compounds may not necessarily involve rupture of this interface, and other conformational changes with the LBD layer may be involved in rendering the LBD layer in the apparently inactive conformation observed in the present structures. However, additional studies with robustly active receptor constructs will help to resolve these questions.

In the GluN1 subunit, the ATD $\alpha 5$ helix C terminus, which harbours exon 5 in an alternatively spliced form of the gene⁴³, in combination with the $\alpha 4$ – $\beta 7$ loop, resides close to the LBD dimer interface, near the C terminus of helix J, and in a plausible position to perturb the conformation of the LBD layer (Fig. 4c). The GluN2B $\alpha 4$ helix C terminus, along with the loop connecting $\alpha 4$ to $\beta 7$, a region implicated in regulation of the NMDA

receptor by polyamines⁴⁴, rest on top of the F and G helices of the GluN1 LBD and close to residues in loop 1 of the GluN2B LBD (Fig. 4d). Thus, although the linking peptides connecting the ATDs to the LBDs have an important role in the transduction of conformational changes between the two layers⁴⁵, direct contacts that harness the predicted large-scale motions of the ATDs³⁶ also play a central role in transmitting changes to the transmembrane, ion channel domain.

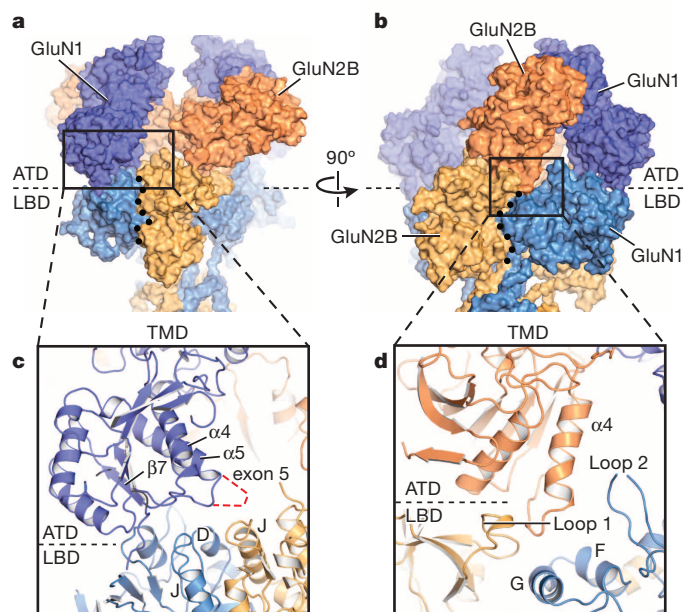


Figure 4 | The ATDs participate in extensive contacts with the LBD layer. **a**, **b**, Surface representation of the ATD and LBD domains, illustrating how the R2 lobe of the GluN1 subunit is poised above its cognate GluN1 LBD and also near the D1–D1 LBD dimer interface (**a**) and how the R2 lobe of the GluN2B subunit participates in contacts with its cognate GluN2B LBD, near an inter LBD dimer interface (**b**). **c**, **d**, Close-up views of potential interactions between the GluN1 R2 lobe and the GluN1 LBD (**c**) and between the GluN2B R2 lobe with regions on its GluN2B LBD (**d**). The GluN2B R2 lobe is also near helices G and F and loop 2 of the GluN1 LBD. In **a** and **b**, the black dots define the approximate intra- and inter-dimer LBD interfaces, respectively. Structure 1 is shown in all panels.

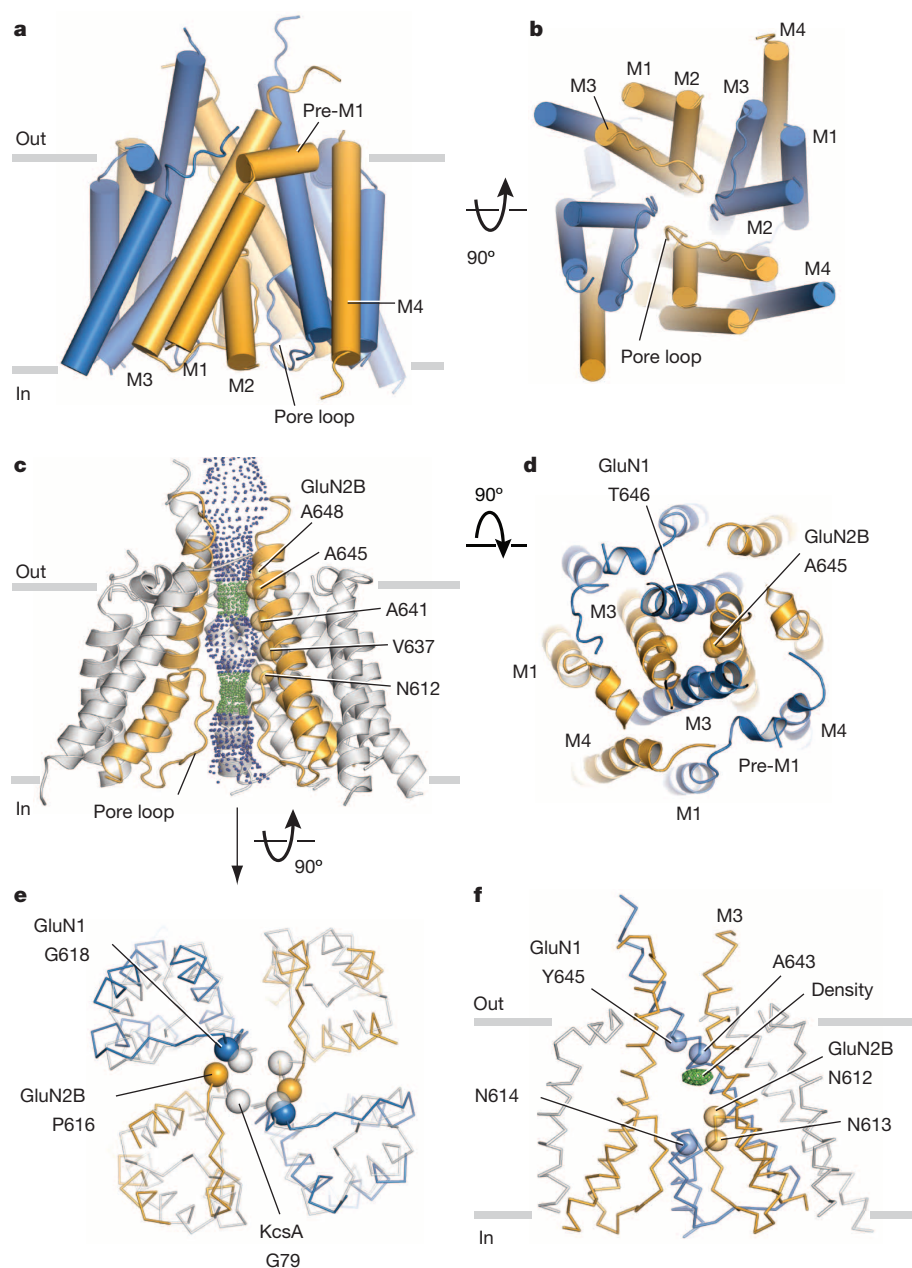


Figure 5 | Transmembrane domain architecture, symmetry and coupling to LBD. **a**, View of the TMD parallel to the membrane. GluN1 subunits are blue and the GluN2B subunits are orange. **b**, View of the TMD, along the pore axis, from the cytoplasmic side of the membrane. **c**, View of a solvent accessible surface carved along the pore axis using the computer program HOLE, parallel to the membrane, showing that the M3 bundle crossing near the extracellular side of the membrane and the entry into the selectivity filter region, from the central aqueous vestibule, form constrictions in the pore. Green dots indicate a pore radius of 1.15–2.3 Å and blue dots define a pore radius greater than 2.3 Å. Because a number of side chains are not included in the structure, due to the moderate resolution of the diffraction data, the size of the pore is approximate. **d**, View of the extracellular ends of the M3 helices of the NMDA receptor. We have highlighted as spheres the α -carbon atoms for residues Thr 646 and Ala 645 in the GluN1–GluN2B structure, respectively. The distances between neighbouring atoms are 6.2, 8.0, 5.4 and 7.1 Å, starting from the α -carbon of GluN2B on the left and going clockwise. **e**, View of the intracellular ends of the TMD of the NMDA receptor in comparison with KcsA. Here the M2 helices of the NMDA receptor were superimposed on the corresponding helices in KcsA, showing the deviation from fourfold symmetry. **f**, Side view of the TMD showing a positive electron density feature (green mesh; 'density') in the central vestibule, calculated using $F_o - F_c$ coefficients and phases from the refined structure. The map is contoured at 2.8 σ . Data set 2 and structure 2 were used in all panels (Extended Data Table 2).

Transmembrane domain

NMDA receptors are calcium permeable and toxic to cells upon over-expression and therefore we introduced mutations known in AMPA receptors to increase receptor desensitization⁴⁶, finding that in the context of the GluN1–GluN2B NMDA receptor, they lead to decreased current amplitudes and enhanced stability of the receptor in detergent micelles. Because native GluN1–GluN2B receptors have a low open probability (P_o)⁴⁷ and the modifications we have introduced further reduce P_o , the functional state of the receptor should be an Ro25-6981, partial agonist bound, closed-blocked channel state.

The electron density associated with data set 2/structure 2 allowed us to position the polypeptide main chain for the M1–M4 helices of all subunits (Supplementary Video 1). To trace the polypeptide associated with the pore loop, we exploited the continuous electron density for this region in the GluN2B subunit D and, by applying non-crystallographic symmetry defined by the transmembrane segments of the other subunits, we traced the three remaining pore loops (Fig. 5a, b). The arrangement of transmembrane helices is like that of the GluA2 AMPA receptor²⁸ (Extended Data Fig. 7a), although in the NMDA receptor we have a more

complete representation of the ion channel pore and putative selectivity filter. The pre-M1 region of the NMDA receptor forms a 'collar' around the extracellular regions of the M3 helices, residing near the boundary of the extracellular side of the membrane. The M1 helix descends across the membrane and makes interactions with the pore-lining M3 helix of the same subunit and the M4 helix of a neighbour. Electron density for the cytoplasmic loop connecting M1 to M2 is weak or missing, and thus this region is absent from the structure. We can visualize the M2 pore helix and most of the extended region of the pore loop forming the selectivity filter and its connection to the N terminus of M3.

The conformation of the polypeptide chain throughout the M2 helix and the pore loop are reminiscent of a potassium channel⁴⁸, although there are differences in the local conformation of the NMDA pore loops, perhaps due to non-glycine residues, SVP, within the canonical GYG motif of potassium channels in the GluN2B subunit (Extended Data Fig. 7b, c) or to the lower resolution of the present structures. Asn residues implicated in voltage-dependent magnesium block⁴⁹ are situated at the turn between the end of M2 and the beginning of the extended filter sequence, in a position to project their side chains into an aqueous vestibule (Fig. 5c, f). Forming

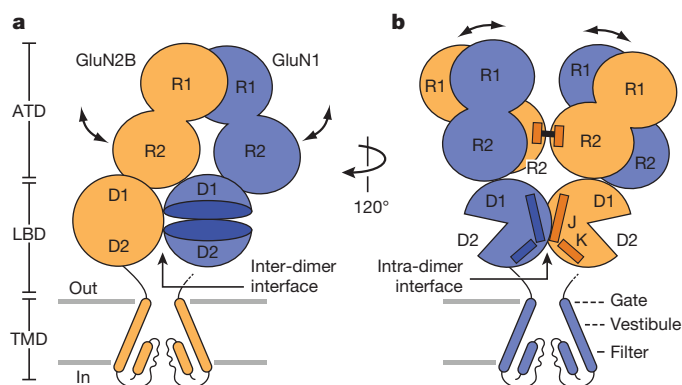


Figure 6 | Schematic of the NMDA receptor. **a**, Shown is a single ATD heterodimer, two LBD clamshells residing in different LBD heterodimers, and the TMD of GluN2B subunits, emphasizing only the M2, pore loop and M3 elements. The line connecting the M3 helix on the right is 'broken' to illustrate that it is connected to the GluN2B LBD behind the visible GluN1 LBD. Double-headed arrows suggest possible movements of ATDs within an ATD heterodimer. **b**, Rotation of the receptor schematic shown in panel **a** by $\sim 120^\circ$ showing two ATD heterodimers, a single LBD heterodimer and the TMD of GluN1 subunits. Double-headed arrows show conformational movements between ATD heterodimers observed in the structures described here. The $\alpha 5$ helices, harbouring the K216C crosslink, are shown as rectangles at the R2–R2 interface. In both schematics, we emphasize how the R2 lobes of the ATDs are positioned such they could modulate inter- and intradimer LBD interfaces and, in turn, the ion channel gate.

the core of the ion channel is the M3 segment, in a similar conformation as the M3 segment in the GluA2 receptor²⁸ (Extended Data Fig. 7a). The extracellular ends of the M3 segments adopt a pyramid-like shape, forming a physical constriction to the ion channel permeation pathway (Fig. 5c, d). The M4 segment resides on the periphery of the transmembrane domain, interacting primarily with the M1 and M3 helices of a neighbouring subunit and extending for several more turns into the cytoplasmic-space than that seen in the GluA2 receptor.

Ion channel gate and central vestibule

The solvent accessible pathway through the ion channel pore from the extracellular side of the membrane to the cytoplasm shows an occlusion near the predicted extracellular boundary of the membrane bilayer, a central vestibule, and a second occlusion at the beginning of the selectivity filter, followed by a solvent accessible pathway to the cytoplasm (Fig. 5c). We speculate that the physical gate of this closed-blocked state of the NMDA receptor is at the bundle crossing of the M3 helices, in a position similar to that of the shut gate of the antagonist-bound GluA2 receptor²⁸. The narrowest constriction is defined by Thr 646 of GluN1 and Ala 645 of GluN2B, within the highly conserved SYTANLAAF motif crucial to ion channel gating, near the extracellular boundary of the M3 helices⁵⁰. Flanking the constriction is a narrow region too small for ion permeation, spanning Val 642–Leu 653 of GluN1 and Ala 641–Ile 652 of GluN2. Although the arrangement of the M3 helices at the ion channel gate diverges from the approximate fourfold symmetry of the AMPA receptor (Fig. 5d)²⁸, we do not yet know if this distinction is inherent to NMDA receptors or is a consequence of their different ligand-bound and functional states.

The pore expands immediately below the M3 bundle crossing to the central vestibule, a cavity flanked by the M3 helices on the sides and the ends of the M2 helices together with the tips of the pore loops on the bottom. A second constriction of the pore is immediately below the central vestibule and is defined by the residues at the beginning of the pore loop (Fig. 5c). Because of the limited resolution of the diffraction data, we are unable to position side-chain groups and to precisely define the location of main chain carbonyl oxygen atoms, and thus are not able to determine if this region of the pore is large enough to conduct ions. Following this constriction, the pore expands to the cytoplasmic space. By contrast with

fourfold symmetric potassium channels, the pore loops of this specific ligand-bound state of the NMDA receptor are arranged with approximate twofold symmetry (Fig. 5e).

In electron density maps derived separately from data set 1 and 2 and their respective structures (Extended Data Table 2), we found positive difference electron density within the central vestibule (Fig. 5f and Extended Data Fig. 7d–f). Because we crystallized the receptor in the presence of 1 mM MK-801, we suggest this electron density feature could represent a trapped MK-801 molecule that occupies multiple positions or orientations within the central vestibule. Unfortunately, we have been unable to validate MK-801 binding by direct binding assays or by electrophysiology experiments, perhaps because of the very low P_o of the receptor construct. Nevertheless, several residues including Ala 643 and Tyr 645 on the GluN1 M3 helix, and Asn 612 (GluN2B) and Asn 613 (GluN2B) on the pore loop 'tips', are near the electron density feature and have been implicated in MK-801 binding¹⁷. We speculate that ion channel blockers, such as MK-801, occupy the central vestibule and block ion conduction by preventing ions from entering the pore loop-lined selectivity filter.

Coupling of ligand binding and transmembrane domains

The coupling of the LBD to the TMD is similar, in principle, to the AMPA receptor²⁸, with the crucial M3 connections to the LBDs proximal or distal to the overall twofold axis for the GluN1 A–C subunits or the GluN2B B–D subunits, respectively (Extended Data Fig. 8). However, the relative orientation of the NMDA receptor TMD with respect to the LBD is distinct from the GluA2 receptor. Specifically, the LBD layer of the NMDA receptor is rotated by $\sim 35^\circ$ around an axis that is approximately coincident with the overall twofold axis of the receptor. At this juncture we do not know if this difference is due to inherent differences between NMDA and AMPA receptors or because this specific NMDA receptor complex is trapped in an ATD antagonist-bound, LBD partial-agonist bound, closed-blocked state.

Conclusion

The GluN1–GluN2B structure harbours an overall twofold symmetry, a layered dimer-of-dimers arrangement of subunits and a positioning of NR1 and NR2B subunits in the A–C and B–D positions defined by the full-length GluA2 receptor²⁸ (Supplementary Video 1). The rich interdigitations and covalent linkage of the R2 lobes of the ATDs to the LBDs provide molecular routes for transmission of allosteric signals to the glycine and glutamate-binding LBD layer, which is organized as a ring of heterodimeric units above the transmembrane ion channel (Fig. 6). Despite closure of the LBD clamshells around partial agonists and an intact 'non-desensitized' D1–D1 interface of the LBD heterodimers, the ion channel gate is in a closed-blocked state, providing the first insights into the structural basis for allosteric inactivation of a NMDA receptor and suggesting that plasticity of the LBD layer may provide a mechanism for modulation of receptor gating. Departing from the fourfold symmetry of the GluA2 transmembrane domain, the pore loops of the NMDA receptor are approximately twofold symmetric. The allosteric antagonist-bound GluN1–GluN2B structure provides a molecular blueprint for the development of new therapeutic agents and a structural framework for biophysical mechanisms of allosteric modulation, gating and ion channel function, as well as a springboard for future studies directed towards determining structures of resting, open and desensitized states and defining locations of ion and blocker binding sites.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 May; accepted 4 June 2014.

Published online 22 June 2014.

1. Traynelis, S. F. *et al.* Glutamate receptor ion channels: structure, regulation, and function. *Pharmacol. Rev.* **62**, 405–496 (2010).

2. Paoletti, P., Bellone, C. & Zhou, Q. NMDA receptor subunit diversity: impact on receptor properties, synaptic plasticity and disease. *Nature Rev. Neurosci.* **14**, 383–400 (2013).
3. Bliss, T. V. P. & Collingridge, G. L. A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* **361**, 31–39 (1993).
4. Soto, D., Altafaj, X., Sindreu, C. & Bayes, A. Glutamate receptor mutations in psychiatric and neurodevelopmental disorders. *Commun. Integr. Biol.* **7**, e27887 (2014).
5. Peery, H. E. et al. Anti-NMDA receptor encephalitis. The disorder, the diagnosis and the immunobiology. *Autoimmun. Rev.* **11**, 863–872 (2012).
6. Keinänen, K. et al. A family of AMPA-selective glutamate receptors. *Science* **249**, 556–560 (1990).
7. Bettler, B. et al. Cloning of a novel glutamate receptor subunit, GluR5: expression in the nervous system during development. *Neuron* **5**, 583–595 (1990).
8. Werner, P., Voigt, M., Keinänen, K., Wisden, W. & Seeburg, P. H. Cloning of a putative high-affinity kainate receptor expressed predominantly in hippocampal CA3 cells. *Nature* **351**, 742–744 (1991).
9. Johnson, J. W. & Ascher, P. Glycine potentiates the NMDA response in cultured mouse brain neurons. *Nature* **325**, 529–531 (1987).
10. Mayer, M. L., Westbrook, G. L. & Guthrie, P. B. Voltage-dependent block by Mg^{2+} of NMDA responses in spinal cord neurones. *Nature* **309**, 261–263 (1984).
11. Nowak, L., Bregestovski, P., Ascher, P., Herbet, A. & Prochiantz, A. Magnesium gates glutamate-activated channels in mouse central neurones. *Nature* **307**, 462–465 (1984).
12. Mayer, M. L. & Westbrook, G. L. Permeation and block of *N*-methyl-D-aspartic acid receptor channels by divalent cations in mouse cultured central neurones. *J. Physiol. (Lond.)* **394**, 501–527 (1987).
13. Moriyoshi, K. et al. Molecular cloning and characterization of the rat NMDA receptor. *Nature* **354**, 31–37 (1991).
14. Monyer, H. et al. Heteromeric NMDA receptors: molecular and functional distinction of subtypes. *Science* **256**, 1217–1221 (1992).
15. Tovar, K. R., McGinley, M. J. & Westbrook, G. L. Triheteromeric NMDA receptors at hippocampal synapses. *J. Neurosci.* **33**, 9150–9160 (2013).
16. Hansen, K. B., Furukawa, H. & Traynelis, S. F. Control of assembly and function of glutamate receptors by the amino terminal domain. *Mol. Pharmacol.* **78**, 535–549 (2010).
17. Kashiwagi, K. et al. Channel blockers acting at *N*-methyl-D-aspartate receptors: differential effects of mutations in the vestibule and ion channel pore. *Mol. Pharmacol.* **61**, 533–545 (2002).
18. Sun, Y. et al. Mechanism of glutamate receptor desensitization. *Nature* **417**, 245–253 (2002).
19. Mayer, M. L. Emerging models of glutamate receptor ion channel structure and function. *Structure* **19**, 1370–1380 (2011).
20. Pøhlsgaard, J., Frydenvang, K., Madsen, U. & Kastrup, J. S. Lessons from more than 80 structures of the GluA2 ligand-binding domain in complex with agonists, antagonists and allosteric modulators. *Neuropharmacology* **60**, 135–150 (2011).
21. Jin, R. et al. Crystal structure and association behaviour of the GluR2 amino-terminal domain. *EMBO J.* **28**, 1812–1823 (2009).
22. Kumar, J., Schuck, P., Jin, R. & Mayer, M. L. The N-terminal domain of GluR6-subtype glutamate receptor ion channels. *Nature Struct. Mol. Biol.* **16**, 631–638 (2009).
23. Karakas, E., Simorowski, N. & Furukawa, H. Structure of the zinc-bound amino-terminal domain of the NMDA receptor NR2B subunit. *EMBO J.* **28**, 3910–3920 (2009).
24. Karakas, E., Simorowski, N. & Furukawa, H. Subunit arrangement and phenylethanolamine binding in GluN1/GluN2 NMDA receptors. *Nature* **475**, 249–253 (2011).
25. Fischer, G. et al. Ro 25–6981, a highly potent and selective blocker of *N*-methyl-D-aspartate receptors containing the NR2B subunit. Characterization *in vitro*. *J. Pharmacol. Exp. Ther.* **283**, 1285–1292 (1997).
26. Watson, G. B. & Lanthorn, T. H. Pharmacological characteristics of cyclic homologues of glycine at the *N*-methyl-D-aspartate receptor-associated glycine site. *Neuropharmacology* **29**, 727–730 (1990).
27. Allan, R. D. et al. Synthesis and activity of a potent *N*-methyl-D-aspartic acid agonist, trans-1-aminocyclobutane-1,3-dicarboxylic acid, and related phosphonic and carboxylic acids. *J. Med. Chem.* **33**, 2905–2915 (1990).
28. Sobolevsky, A. I., Rosconi, M. P. & Gouaux, E. X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor. *Nature* **462**, 745–756 (2009).
29. Salussolia, C. L., Prodromou, M. L., Borker, P. & Wollmuth, L. P. Arrangement of subunits in functional NMDA receptors. *J. Neurosci.* **31**, 11295–11304 (2011).
30. Riou, M., Stroebel, D., Edwardson, J. M. & Paoletti, P. An alternating GluN1–2–1–2 subunit arrangement in mature NMDA receptors. *PLoS ONE* **7**, e35134 (2012).
31. Lee, C. H. & Gouaux, E. Amino terminal domains of the NMDA receptor are organized as local heterodimers. *PLoS ONE* **6**, e19180 (2011).
32. Furukawa, H., Singh, S., Mancusso, R. & Gouaux, E. Subunit arrangement and function in NMDA receptors. *Nature* **438**, 185–192 (2005).
33. Mony, L., Kew, J. N., Gunthrope, M. J. & Paoletti, P. Allosteric modulators of NR2B-containing NMDA receptors: molecular mechanisms and therapeutic potential. *Br. J. Pharmacol.* **157**, 1301–1317 (2009).
34. Reichling, D. B. & MacDermott, A. B. Lanthanum actions on excitatory amino acid-gated currents and voltage-gated calcium currents in rat dorsal horn neurones. *J. Physiol. (Lond.)* **441**, 199–218 (1991).
35. Sherry, A. D., Newman, A. D. & Gutz, C. G. The activation of concavalin A by lanthanide ions. *Biochemistry* **14**, 2191–2196 (1975).
36. Zhu, S., Stroebel, D., Yao, C. A., Taly, A. & Paoletti, P. Allosteric signaling and dynamics of the clamshell-like NMDA receptor GluN1 N-terminal domain. *Nature Struct. Mol. Biol.* **20**, 477–485 (2013).
37. Weston, M. C., Schuck, P., Ghosal, A., Rosenmund, C. & Mayer, M. L. Conformational restriction blocks glutamate receptor desensitization. *Nature Struct. Mol. Biol.* **13**, 1120–1127 (2006).
38. Gielen, M. et al. Structural rearrangements of NR1/NR2A NMDA receptors during allosteric inhibition. *Neuron* **57**, 80–93 (2008).
39. Hansen, K. B., Ogden, K. K. & Traynelis, S. F. Subunit-selective allosteric inhibition of glycine binding to NMDA receptors. *J. Neurosci.* **32**, 6197–6208 (2012).
40. Inanobe, A., Furukawa, H. & Gouaux, E. Mechanism of partial agonist action at the NR1 subunit of NMDA receptors. *Neuron* **47**, 71–84 (2005).
41. Erreger, K. et al. Subunit-specific agonist activity at NR2A-, NR2B-, NR2C-, and NR2D-containing *N*-methyl-D-aspartate glutamate receptors. *Mol. Pharmacol.* **72**, 907–920 (2007).
42. Armstrong, N. & Gouaux, E. Mechanisms for activation and antagonism of an AMPA-sensitive glutamate receptor: crystal structures of the GluR2 ligand binding core. *Neuron* **28**, 165–181 (2000).
43. Sugihara, H., Moriyoshi, K., Ishii, T., Masu, M. & Nakanishi, S. Structures and properties of seven isoforms of the NMDA receptor generated by alternative splicing. *Biochem. Biophys. Res. Commun.* **185**, 826–832 (1992).
44. Mony, L., Zhu, S., Carvalho, S. & Paoletti, P. Molecular basis of positive allosteric modulation of GluN2B NMDA receptors by polyamines. *EMBO J.* **30**, 3134–3146 (2011).
45. Gielen, M., Siegler Retchless, B., Mony, L., Johnson, J. W. & Paoletti, P. Mechanism of differential control of NMDA receptor activity by NR2 subunits. *Nature* **459**, 703–707 (2009).
46. Yelshansky, M. V., Sobolevsky, A. I., Jatzke, C. & Wollmuth, L. P. Block of AMPA receptor desensitization by a point mutation outside the ligand-binding domain. *J. Neurosci.* **24**, 4728–4736 (2004).
47. Erreger, K., Dravid, S. M., Banke, T. G., Wyllie, D. J. & Traynelis, S. F. Subunit-specific gating controls rat NR1/NR2A and NR1/NR2B NMDA channel kinetics and synaptic signalling profiles. *J. Physiol. (Lond.)* **563**, 345–358 (2005).
48. Doyle, D. A. et al. The structure of the potassium channel: molecular basis of K^{+} conduction and selectivity. *Science* **280**, 69–77 (1998).
49. Burnashev, N. et al. Control by asparagine residues of calcium permeability and magnesium blockade in the NMDA receptor. *Science* **257**, 1415–1419 (1992).
50. Kuner, T., Seeburg, P. H. & Guy, H. R. A common architecture for K^{+} channels and ionotropic glutamate receptors. *Trends Neurosci.* **26**, 27–32 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements All members of the Gouaux laboratory are gratefully acknowledged for their support and assistance, especially L. Chen, K. Duerr and K. Wang. We thank L. Vaskalis for assistance with the figures, G. Westbrook and C. Jahr for comments on the manuscript, H. Owen for proofreading and I. Baconguis for making the animation. E.G. acknowledges the generous support of R. LaCrute, B. LaCrute and J. LaCrute. This work was also supported by an Oregon Brain Institute Graduate Student Fellowship (C.H.L.), the NIH (E.G.) and the Vollum Institute (E.G.). E.G. is an investigator of the Howard Hughes Medical Institute.

Author Contributions C.-H.L. contributed to all aspects of the research; W.L. carried out crystallographic analysis; J.C.M. carried out molecular biology, cell culture, electrophysiology and ligand binding experiments; A.G. performed molecular biology, cell culture, receptor purification and crystallization studies; J.D. and X.S. analysed the structures; E.G. directed the research; and all authors contributed to the preparation of the manuscript.

Author Information The coordinates and structure factors for the structure have been deposited in the Protein Data Bank under accession code 4TLT and 4TLM for structure 1 and structure 2, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.G. (gouauxe@ohsu.edu).

Non-local propagation of correlations in quantum systems with long-range interactions

Philip Richerme¹, Zhe-Xuan Gong¹, Aaron Lee¹, Crystal Senko¹, Jacob Smith¹, Michael Foss-Feig¹, Spyridon Michalakis², Alexey V. Gorshkov¹ & Christopher Monroe¹

The maximum speed with which information can propagate in a quantum many-body system directly affects how quickly disparate parts of the system can become correlated^{1–4} and how difficult the system will be to describe numerically⁵. For systems with only short-range interactions, Lieb and Robinson derived a constant-velocity bound that limits correlations to within a linear effective ‘light cone’⁶. However, little is known about the propagation speed in systems with long-range interactions, because analytic solutions rarely exist and because the best long-range bound⁷ is too loose to accurately describe the relevant dynamical timescales for any known spin model. Here we apply a variable-range Ising spin chain Hamiltonian and a variable-range XY spin chain Hamiltonian to a far-from-equilibrium quantum many-body system and observe its time evolution. For several different interaction ranges, we determine the spatial and time-dependent correlations, extract the shape of the light cone and measure the velocity with which correlations propagate through the system. This work opens the possibility for studying a wide range of many-body dynamics in quantum systems that are otherwise intractable.

Lieb–Robinson bounds⁶ have strongly influenced our understanding of locally interacting, quantum many-body systems. They restrict the many-body dynamics to a well-defined causal region outside of which correlations are exponentially suppressed⁸, analogous to causal light cones that arise in relativistic theories. These bounds have enabled a number of important proofs in condensed-matter physics^{5,7,9–11}, and also constrain the timescales on which quantum systems might thermalize^{12–14} and the maximum speed that information can be sent through a quantum channel¹⁵. Recent experimental work has observed linear (that is, Lieb–Robinson-like) correlation growth over six sites in a one-dimensional quantum gas¹⁶.

When interactions in a quantum system are long range, the speed with which correlations build up between distant particles is no longer guaranteed to obey the Lieb–Robinson prediction. Indeed, for sufficiently long-range interactions, the notion of locality is expected to break down completely¹⁷. Inapplicability of the Lieb–Robinson bound means that comparatively little can be predicted about the growth and propagation of correlations in long-range-interacting systems, although there have been several recent theoretical and numerical advances^{2,3,7,17–20}.

Here we report direct measurements of the shape of the causal region and the speed at which correlations propagate in an Ising spin chain and a newly implemented XY spin chain. The experiment is effectively decoherence free and serves as an initial probe of the many-body dynamics of isolated quantum systems. Within this broad experimental framework, studies of entanglement growth²¹, thermalization^{12,14} or other dynamical processes—with or without controlled decoherence—can be realized. Scaling such quantum simulations to larger system sizes is straightforward (Methods), unlike ground-state or equilibrium studies that typically must consider diabatic effects^{22,23}.

To induce the spread of correlations, we perform a global quench by suddenly switching on the spin–spin couplings across the entire chain and allowing the system to evolve coherently. The dynamics following a global quench can be highly non-intuitive; one picture is that entangled quasiparticles created at each site propagate outwards, correlating distant parts

of the system through multiple interference pathways¹³. This process differs substantially from local quenches²¹, where a single site emits quasiparticles that may travel ballistically^{3,13}, resulting in a different causal region and propagation speed than in a global quench (even for the same spin model).

The effective spin-1/2 system is encoded into the $^2S_{1/2}|F=0, m_F=0\rangle$ and $|F=1, m_F=0\rangle$ hyperfine ‘clock’ states of trapped $^{171}\text{Yb}^+$ ions, denoted $|\downarrow\rangle_z$ and $|\uparrow\rangle_z$, respectively²⁴. We initialize a chain of 11 ions by optically pumping to the product state $|\downarrow\downarrow\downarrow\dots\rangle_z$ (Fig. 1). At time $t=0$, we quench the system by applying phonon-mediated, laser-induced forces^{25–27} to yield an Ising or XY model Hamiltonian (Methods)

$$H_{\text{Ising}} = \sum_{i < j} J_{ij} \sigma_i^x \sigma_j^x \quad (1)$$

$$H_{\text{XY}} = \frac{1}{2} \sum_{i < j} J_{ij} (\sigma_i^x \sigma_j^x + \sigma_i^z \sigma_j^z) \quad (2)$$

where σ_i^γ ($\gamma = x, y, z$) is the Pauli matrix acting on the i th spin, J_{ij} (in cyclic frequency) is the coupling strength between spins i and j , and we use units in which Planck’s constant equals 1. For both model Hamiltonians, the

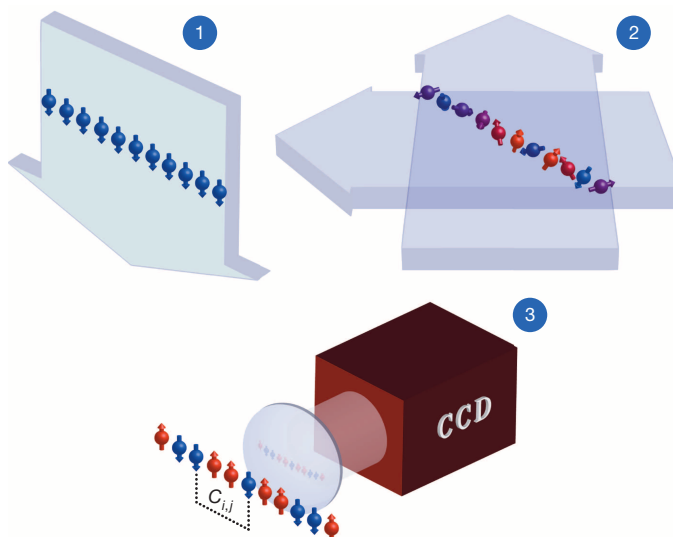


Figure 1 | Sketch of experimental protocol. Step (1): the experiment is initialized by optically pumping all 11 spins to the state $|\downarrow\rangle_z$. Step (2): after initialization, the system is quenched by applying laser-induced forces on the ions, yielding an effective Ising or XY spin chain (see text for details). Step (3): after allowing dynamical evolution of the system, the projection of each spin along the \hat{z} direction is imaged onto a charge-coupled device (CCD) camera. Such measurements allow us to construct any possible correlation function $C_{i,j}$ along \hat{z} .

¹Joint Quantum Institute, University of Maryland Department of Physics and National Institute of Standards and Technology, College Park, Maryland 20742, USA. ²Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, California 91125, USA.

spin–spin interaction matrix J_{ij} contains tunable, long-range couplings that fall off approximately algebraically as $J_{ij} \propto 1/|i-j|^\alpha$ (ref. 26). We vary the interaction range α by adjusting a combination of trap and laser parameters²² (Methods), choosing $\alpha \approx 0.63, 0.83, 1.00$ or 1.19 for these experiments.

After quenching to the Ising or XY model with our chosen value of α , we allow coherent evolution for various lengths of time before resolving the spin state of each ion using a charge-coupled device camera. The experiments at each time step are repeated 4,000 times to collect statistics. To observe the build-up of correlations, we use the measured spin states to construct the connected correlation function

$$C_{ij}(t) = \langle \sigma_i^z(t) \sigma_j^z(t) \rangle - \langle \sigma_i^z(t) \rangle \langle \sigma_j^z(t) \rangle \quad (3)$$

between any pair of ions at any time. Because the system is initially in a product state, $C_{ij}(0) = 0$ everywhere. As the system evolves away from a product state, evaluating equation (3) at all points in space and time provides the shape of the light-cone boundary and the correlation propagation velocity for our long-range spin models.

Figure 2 shows the results of globally quenching the system to a long-range Ising model for four different interaction ranges. In each case, we extract the light-cone boundary by measuring the time it takes a correlation of fixed amplitude (here $C_{ij} = 0.04 \approx 0.1 C_{ij}^{\max}$, where C_{ij}^{\max} is the largest connected correlation between two ions) to travel an ion–ion separation distance r . For strongly long-range interactions ($\alpha \leq 1$), we observe accelerating information transfer through the chain. This fast propagation of correlations is not surprising, because even the direct long-range coupling between distant spins produces correlations in a time $t \propto 1/J_{ij} \approx r^\alpha$. However, increasing propagation velocities quickly surpass the Lieb–Robinson velocity for a system with equivalent nearest-neighbour-only interactions, $v_{\text{LR}} = 12eJ_{\text{max}}$ where e is Euler’s number and J_{max} is the maximum Ising coupling strength for a given spin–spin coupling matrix (Fig. 2c, f, i). This serves as experimental confirmation that predictions based on the Lieb–Robinson result—including those that bound the growth of entanglement or set thermalization timescales—are no longer applicable when interactions are sufficiently long range.

For the specific case of the pure Ising model, the correlations at any time can be predicted by an exact analytic solution^{18,28}:

$$C_{ij}(t) = \frac{1}{2} \prod_{k \neq i,j} \cos[2(J_{i,k} + J_{j,k})t] + \frac{1}{2} \prod_{k \neq i,j} \cos[2(J_{i,k} - J_{j,k})t] - \prod_{k \neq i} \cos[2J_{i,k}t] \prod_{k \neq j} \cos[2J_{j,k}t] \quad (4)$$

In equation (4), correlations can only build up between sites i and j that are coupled either directly or through a single intermediate spin k ; processes which couple through more than one intermediate site are prohibited. For instance, if the J_{ij} couplings are nearest-neighbour-only then $C_{ij}(t) = 0$ for all $|i-j| > 2$. This property holds for any commuting Hamiltonian (Methods) and explains why the spatial correlations shown in Fig. 2 become weaker for shorter-range systems.

The products of cosines in equation (4) with many different oscillation frequencies result in the observed decay of correlations when $t \gtrsim 0.1/J_{\text{max}}$. At later times, rephasing of these oscillations creates revivals in the spin–spin correlation. One such partial revival occurs at $t = 2.44/J_{\text{max}}$ for $\alpha = 0.63$ (Extended Data Fig. 1), verifying that our system remains coherent on a timescale much longer than that which determines the light-cone boundary.

We repeat the quench experiments for an XY model Hamiltonian using the same set of interaction ranges α (Fig. 3). Dynamical evolution and the spread of correlations in long-range-interacting XY models are much more complex than in the Ising case because the Hamiltonian contains non-commuting terms. As a result, there exists no exact analytic solution comparable to equation (4).

Compared with the correlations observed for the Ising Hamiltonian, correlations in the XY model are much stronger at longer distances (for example, compare Fig. 2j with Fig. 3j). Processes coupling through multiple intermediate sites (which were disallowed in the commuting Ising Hamiltonian) now have a critical role in building correlations between distant spins. These processes may also explain our observation of a steeper

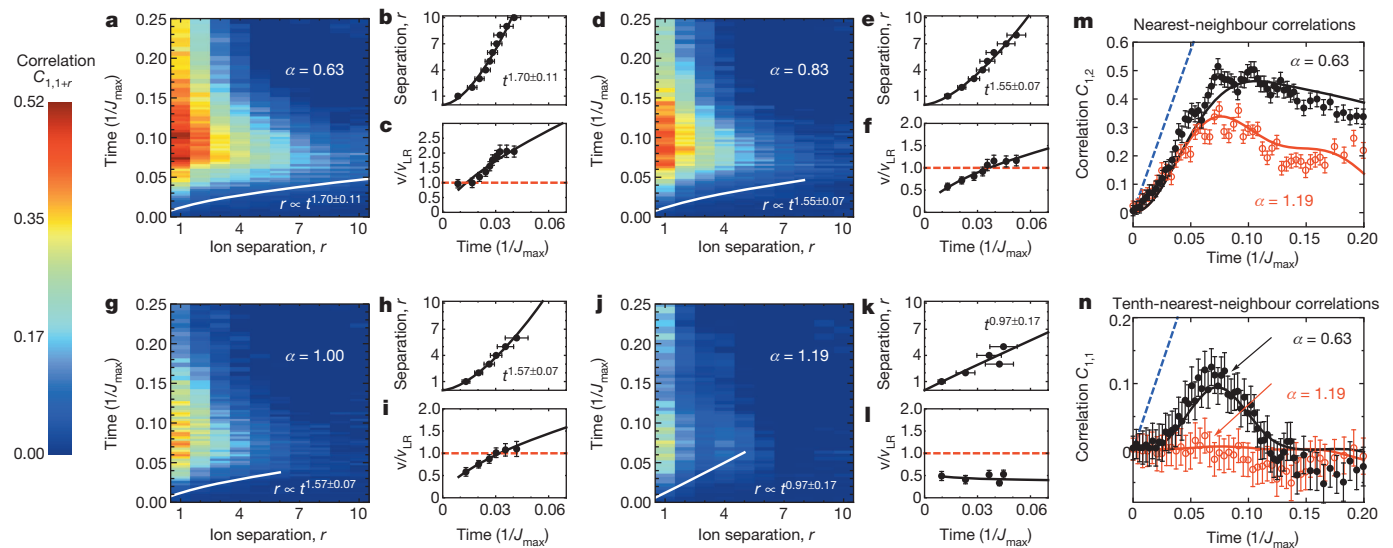


Figure 2 | Measured quench dynamics in a long-range Ising model. **a–c**, Spatial and time-dependent correlations (**a**), extracted light-cone boundary (**b**) and correlation propagation velocity (**c**) following a global quench of a long-range Ising model with $\alpha = 0.63$. The curvature of the boundary shows an increasing propagation velocity (**b**), quickly exceeding the short-range Lieb–Robinson velocity bound, v_{LR} (**c**). Solid lines give a power-law fit to the data, which slightly depends on the choice of fixed contour C_{ij} . **d–f**, Complementary plots for $\alpha = 0.83$ (**d–f**), $\alpha = 1.00$ (**g–i**) and $\alpha = 1.19$ (**j–l**). As the range of the interactions decreases, correlations do not

propagate as far or as quickly through the chain; the short-range velocity bound v_{LR} is not exceeded for our shortest-range interaction. **m, n**, Nearest-neighbour (**m**) and tenth-nearest-neighbour (**n**) correlations for our shortest- and longest-range interactions show excellent agreement with the decoherence-free exact solution (with no adjustable parameters) from equation (4) (solid). The dashed blue curves show an improved long-range bound valid for any commuting Hamiltonian (Methods). Error bars, 1 s.d.

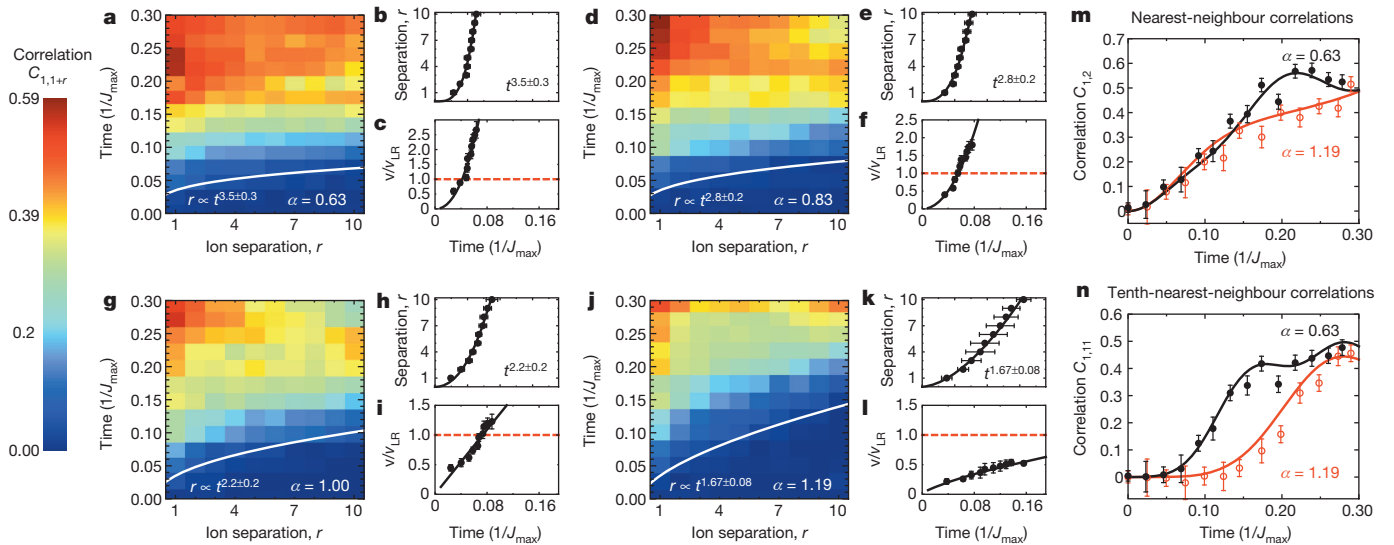


Figure 3 | Measured quench dynamics in a long-range XY model. Global quench of a long-range XY model with four different interaction ranges. **a–l**, Panel descriptions match those in Fig. 2. In each case, when compared with the Ising model, correlations between distant sites in the XY model are stronger and build up more quickly. For the shortest-range interaction (**j–l**), we observe a faster-than-linear growth of the light-cone boundary,

power-law scaling of the light-cone boundary in the XY model. However, without an exact solution there is no a-priori reason to assume a power-law light-cone edge (used for the fits in Fig. 3); deviations from power-law behaviour might reveal themselves for larger system sizes.

An important observation in Fig. 3j–l is that of faster-than-linear light-cone growth for our shortest-range interaction, with $\alpha = 1.19$. Although faster-than-linear growth is expected for $\alpha < 1$ (see discussion of Ising model), there is no consensus on whether such behaviour is generically expected for $\alpha > 1$. Our experimental observation has prompted us to numerically check the light-cone shape for $\alpha = 1.19$; we find that faster-than-linear scaling persists in systems of up to 22 spins before our calculations break down (Extended Data Fig. 2).

Whether such scaling continues beyond ~ 30 spins is a question that, at present, quantum simulators are best positioned to answer. In Figs 2m, n and 3m, n the excellent agreement between data and theory demonstrates that experiments produce the correct results in a regime still solvable by classical computers. For larger systems, where numerical evolution of the Schrödinger equation fails, the quality of quantum simulations could still be benchmarked against the exact Ising solution of equation (4). Finding close agreement in the Ising case would then build confidence in an XY model simulation, which cannot be validated by any other known method.

For the XY model, we additionally study the spatial decay of correlations outside the light-cone boundary. The data (Fig. 4) is well described by fits to exponentially decaying functions. Recent theoretical work²⁰ predicts an initial decay of spatial correlations bounded by an exponential, followed by a power-law decay; we speculate that much larger system sizes and several hundred thousand repetitions of each data point (to reduce the shot-noise uncertainty sufficiently) would be necessary to see this effect.

A perturbative treatment of time evolution under the XY Hamiltonian yields the short-time approximation for the correlation function $C_{ij}(t) \approx (J_{ij}t)^2$. These values are plotted as dashed lines along with the data in Fig. 4. Although the perturbative result matches the data early on, it fails to describe the dynamics at longer evolution times. The discrepancies indicate that the light-cone shapes observed in the XY model are fundamentally non-perturbative; rather, they result from the build-up of correlations through multiple intermediate sites and cannot be described by any known analytical method.

despite having $\alpha > 1$; no known analytic theory predicts this effect.

m, n, Measured nearest-neighbour and tenth-nearest-neighbour correlations closely match the numerical solution found by evolving the Schrödinger equation of an XY model (equation (2)) with no free parameters and no decoherence. Error bars, 1 s.d.

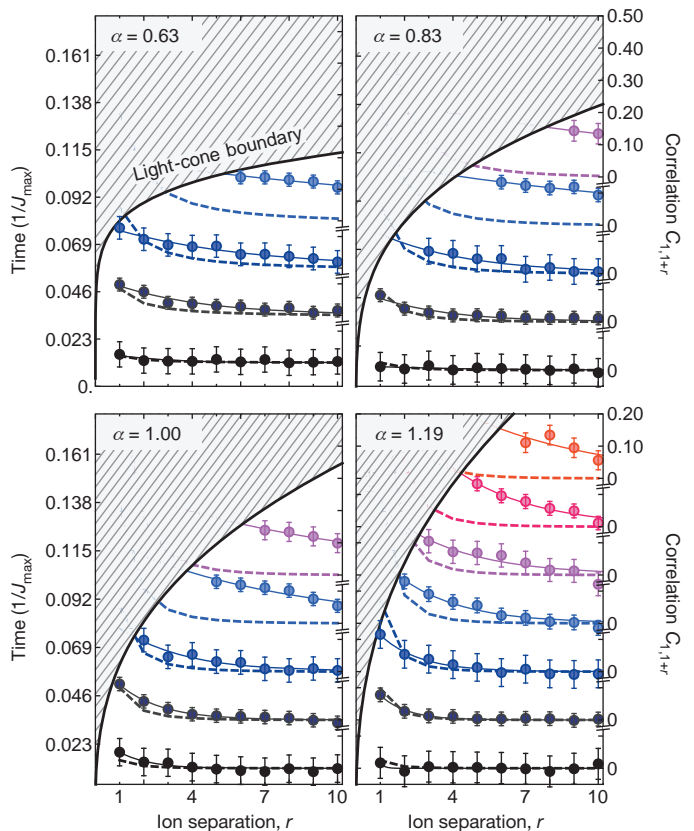


Figure 4 | Correlations and dynamics beyond the perturbative regime.

Decay of spatial correlations outside the light-cone boundaries for a long-range XY model with $\alpha = 0.63, 0.83, 1.00$ or 1.19 . The hatched region indicates the area inside the light-cone boundary $C_{ij} = 0.15$. The data corresponds to times indicated by tickmarks on the left axis. Solid lines give an exponential fit to the data and dashed lines show the predictions from a perturbative calculation. Perturbation theory does not accurately describe the dynamics at later times. Associated data and theoretical results are similarly coloured to guide the eye. Error bars, 1 s.d.

We have presented experimental observations of the causal region and propagation velocities for correlations following global quenches in Ising and XY spin models. The long-range interactions in our system lead to a breakdown of the locality associated with Lieb–Robinson bounds, and dynamical evolution in the XY model leads to results that cannot be described by analytic or perturbative theory. Our work demonstrates that quantum simulators with only a few tens of spins can be an important tool for investigating and enriching our understanding of dynamics in complex many-body systems.

METHODS SUMMARY

We generate spin–spin interactions by applying spin-dependent optical dipole forces to ions confined in a three-layer linear Paul trap with a 4.8 MHz radial frequency. Two off-resonance laser beams with a wavevector difference Δk along a principal axis of transverse motion globally address the ions and drive stimulated Raman transitions. The two beams contain a pair of beat-note frequencies symmetrically detuned from the resonant transition at $\nu_0 = 12.642819$ GHz by a frequency μ , comparable to the transverse motional mode frequencies. In the Lamb–Dicke regime, this results in the Ising-type Hamiltonian in equation (1)^{25,26} with

$$J_{ij} = \Omega^2 \omega_R \sum_{m=1}^N \frac{b_{i,m} b_{j,m}}{\mu^2 - \omega_m^2}$$

where Ω is the global Rabi frequency, $\omega_R = \hbar \Delta k^2 / 2M$ (\hbar , Planck's constant divided by 2π) is the recoil frequency, $b_{i,m}$ is the normal-mode matrix²⁹ and ω_m are the transverse mode frequencies. The coupling profile may be approximated as a power-law decay $J_{ij} \approx J_0 / |i - j|^\alpha$, where in principle α can be tuned between 0 and 3 by varying the laser detuning μ or the trap frequencies ω_m (refs 22, 26).

We implement a tunable-range XY model by adding an effective transverse magnetic field $B \sum_i \sigma_i^y$ to the pure Ising Hamiltonian with an additional laser beat-note frequency at ν_0 . In the limit $B \gg J$, processes governed by the $\sigma_i^x \sigma_j^x$ coupling which flip two spins along y (for example $\sigma^+ \sigma^+$, where here $\sigma^\pm = \sigma^z \pm i\sigma^y$) are energetically forbidden, leaving only the energy-conserving flip-flop terms ($\sigma^+ \sigma^- + \sigma^- \sigma^+$). At times $t = n/B$ (with integer n), the dynamics of the transverse field rephases and leaves only the pure XY Hamiltonian of equation (2).

In the limit $B > \eta_m \Omega$, where $\eta_m = \Delta k \sqrt{\hbar / 2M\omega_m}$, phonon contributions from the large, non-commuting transverse field can lead to unwanted spin–motion entanglement at the end of an experiment³⁰. Therefore, this method of generating an XY model requires the hierarchy $J \ll B \ll \eta_m \Omega$ for all m . For our typical trap parameters, $J_{\max} \approx 400$ Hz, $B \approx 4$ kHz and $\eta_m \Omega \approx 20$ kHz.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 January; accepted 28 April 2014.

- Nachtergaele, B., Ogata, Y. & Sims, R. Propagation of correlations in quantum lattice systems. *J. Stat. Phys.* **124**, 1 (2006).
- Schachenmayer, J., Lanyon, B., Roos, C. & Daley, A. Entanglement growth in quench dynamics with variable range interactions. *Phys. Rev. X* **3**, 031015 (2013).
- Hauke, P. & Tagliacozzo, L. Spread of correlations in long-range interacting quantum systems. *Phys. Rev. Lett.* **111**, 207202 (2013).
- Hazzard, K. R. A., Manmana, S. R., Foss-Feig, M. & Rey, A. M. Far-from-equilibrium quantum magnetism with ultracold polar molecules. *Phys. Rev. Lett.* **110**, 075301 (2013).
- Eisert, J., Cramer, M. & Plenio, M. Area laws for the entanglement entropy. *Rev. Mod. Phys.* **82**, 277 (2010).
- Lieb, E. & Robinson, D. The finite group velocity of quantum spin systems. *Commun. Math. Phys.* **28**, 251–257 (1972).
- Hastings, M. & Kitaev, T. Spectral gap and exponential decay of correlations. *Commun. Math. Phys.* **265**, 781–804 (2006).
- Bravyi, S., Hastings, M. B. & Verstraete, F. Lieb–Robinson bounds and the generation of correlations and topological quantum order. *Phys. Rev. Lett.* **97**, 050401 (2006).

- Nachtergaele, B. & Sims, R. Lieb–Robinson bound and the exponential clustering theorem. *Commun. Math. Phys.* **265**, 119–130 (2006).
- Hastings, M. An area law for one-dimensional quantum systems. *J. Stat. Mech.* **2007**, P08024 (2007).
- Michalakis, S. Stability of the area law for the entropy of entanglement. Preprint at <http://arxiv.org/abs/1206.6900> (2012).
- Rigol, M., Dunjko, V., Yurovsky, V. & Olshanii, M. Relaxation in a completely integrable many-body quantum system: an ab initio study of the dynamics of the highly excited states of 1D lattice hard-core bosons. *Phys. Rev. Lett.* **98**, 050405 (2007).
- Calabrese, P. & Cardy, J. Time dependence of correlation functions following a quantum quench. *Phys. Rev. Lett.* **96**, 136801 (2006).
- Gong, Z.-X. & Duan, L.-M. Prethermalization and dynamic phase transition in an isolated trapped ion spin chain. *New J. Phys.* **15**, 113051 (2013).
- Bose, S. Quantum communication through spin chain dynamics: an introductory overview. *Contemp. Phys.* **48**, 13–30 (2007).
- Cheneau, M. et al. Light-cone-like spreading of correlations in a quantum many-body system. *Nature* **481**, 484–487 (2012).
- Eisert, J., van den Worm, M., Manmana, S. & Kastner, M. Breakdown of quasi-locality in long-range quantum lattice models. *Phys. Rev. Lett.* **111**, 260401 (2013).
- van den Worm, M., Sawyer, B., Bollinger, J. & Kastner, M. Relaxation timescales and decay of correlations in a long-range interacting quantum simulator. *New J. Phys.* **15**, 083007 (2013).
- Jünemann, J., Cadarso, A., Perez-Garcia, D., Bermudez, A. & Garcia-Ripoll, J. J. Lieb–Robinson bounds for spin-boson lattice models and trapped ions. *Phys. Rev. Lett.* **111**, 230404 (2013).
- Gong, Z.-X., Foss-Feig, M., Michalakis, S. & Gorshkov, A. V. Persistence of locality in systems with power-law interactions. Preprint at <http://arxiv.org/abs/1401.6174> (2014).
- Jurcevic, P. et al. Quasiparticle engineering and entanglement propagation in a quantum many-body system. *Nature* <http://dx.doi.org/10.1038/nature13461> (this issue).
- Islam, R. et al. Emergence and frustration of magnetic order with variable-range interactions in a trapped ion quantum simulator. *Science* **340**, 583–587 (2013).
- Richerme, P. et al. Quantum catalysis of magnetic phase transitions in a quantum simulator. *Phys. Rev. Lett.* **111**, 100506 (2013).
- Olmschenk, S. et al. Manipulation and detection of a trapped Yb⁺ hyperfine qubit. *Phys. Rev. A* **76**, 052314 (2007).
- Mølmer, K. & Sørensen, A. Multiparticle entanglement of hot trapped ions. *Phys. Rev. Lett.* **82**, 1835 (1999).
- Porrás, D. & Cirac, J. I. Effective quantum spin systems with trapped ions. *Phys. Rev. Lett.* **92**, 207901 (2004).
- Kim, K. et al. Entanglement and tunable spin-spin couplings between trapped ions using multiple transverse modes. *Phys. Rev. Lett.* **103**, 120502 (2009).
- Foss-Feig, M., Hazzard, K. R. A., Bollinger, J. J. & Rey, A. M. Nonequilibrium dynamics of arbitrary-range Ising models with decoherence: an exact analytic solution. *Phys. Rev. A* **87**, 042101 (2013).
- James, D. F. V. Quantum dynamics of cold trapped ions with application to quantum computation. *Appl. Phys. B* **66**, 181–190 (1998).
- Wang, C.-C. J. & Freericks, J. K. Intrinsic phonon effects on analog quantum simulators with ultracold trapped ions. *Phys. Rev. A* **86**, 032329 (2012).

Acknowledgements We thank J. Preskill, A. M. Rey, K. Hazzard, A. Daley, J. Schachenmayer, M. Kastner, S. Manmana and L.-M. Duan for discussions. This work is supported by the US Army Research Office (ARO) Award W911NF0710576 with funds from the DARPA Optical Lattice Emulator Program, ARO award W911NF0410234 with funds from the IARPA MQCO Program, and the US NSF Physics Frontier Center at JQI. M.F.-F. thanks the NRC for support. S.M. acknowledges funding provided by the Institute for Quantum Information and Matter, an NSF Physics Frontier Center with the support of the Gordon and Betty Moore Foundation (through grant GBMF1250).

Author Contributions Experiments and data analysis were performed by P.R., A.L., C.S., J.S. and C.M. Theoretical calculations were done by Z.-X.G., M.F.-F., S.M., and A.V.G. All authors contributed to the preparation of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.R. (richerme@umd.edu).

Quasiparticle engineering and entanglement propagation in a quantum many-body system

P. Jurcevic^{1,2*}, B. P. Lanyon^{1,2*}, P. Hauke^{1,3}, C. Hempel^{1,2}, P. Zoller^{1,3}, R. Blatt^{1,2} & C. F. Roos^{1,2}

The key to explaining and controlling a range of quantum phenomena is to study how information propagates around many-body systems. Quantum dynamics can be described by particle-like carriers of information that emerge in the collective behaviour of the underlying system, the so-called quasiparticles¹. These elementary excitations are predicted to distribute quantum information in a fashion determined by the system's interactions². Here we report quasiparticle dynamics observed in a quantum many-body system of trapped atomic ions^{3,4}. First, we observe the entanglement distributed by quasiparticles as they trace out light-cone-like wavefronts^{5–11}. Second, using the ability to tune the interaction range in our system, we observe information propagation in an experimental regime where the effective-light-cone picture does not apply^{7,12}. Our results will enable experimental studies of a range of quantum phenomena, including transport^{13,14}, thermalization¹⁵, localization¹⁶ and entanglement growth¹⁷, and represent a first step towards a new quantum-optic regime of engineered quasiparticles with tunable nonlinear interactions.

Experimental study of the quantum coherent properties of quasiparticles represents a significant challenge over a broad range of disciplines, from investigating the propagation of quantum correlations by magnons or phonons in naturally occurring systems, to creating exotic anyons in engineered quantum matter for topological quantum technology¹⁸. Very recently, in an engineered system of 10 to 18 atoms in an optical lattice^{9,10}, the propagation of correlations by magnons was observed. Detecting the quantum nature of the correlations distributed by quasiparticles (that is, entanglement) has not yet been achieved.

We present experiments in an engineered one-dimensional system of atomic ions, whose dynamics are accurately described by a lattice model of interacting spins. The salient features of our system are as follows: first, we are able to inject localized excitations into the system and, in the subsequent dynamics, measure arbitrary multi-particle correlation functions, thereby allowing the possibility of detecting and quantifying entanglement. Second, we can tune the ion–ion interaction range from effectively nearest-neighbour to infinite range. In each case, a new set of quasiparticles is created with unique dynamical properties. In contrast to quantum simulation, where one aims to study computationally complex processes, our goal is to manipulate individual quasiparticles and their properties precisely, and observe the information and quantum correlations that they transport.

In any quantum system with finite-range interactions (for example, nearest-neighbour), quasiparticles have a finite maximal velocity, known as the Lieb–Robinson velocity^{5,19,20}. Such systems exhibit an effective light cone for quantum dynamics that strictly bounds the speed with which information propagates, as recently observed^{9,10}. However, many natural and engineered quantum systems exhibit long-range interactions that decay as a power law, such as dipole–dipole or van der Waals interactions. Understanding the quantum dynamics in this more general case is an active field of theoretical research^{12,17,21,22}. For weakly long-range interactions, the light-cone picture remains a good description, but as the interaction range increases, the maximal velocity in the system is predicted

to diverge and the light-cone picture no longer applies¹². Here, we experimentally access these regimes.

We control the quantum state and interactions of the valence electron in each (⁴⁰Ca⁺) ion, which can be in one of two states $|S_{1/2}, m = +1/2\rangle$ or $|D_{5/2}, m = +5/2\rangle$ that we label $|\downarrow\rangle$ and $|\uparrow\rangle$ respectively. The quasiparticles of our system are therefore collective electronic excitations. In the presence of laser-driven interactions, our system is well described by a one-dimensional model of interacting spins, with Hamiltonian^{3,4}

$$H_{\text{Ising}} = \hbar \sum_{i < j} J_{ij} \sigma_i^x \sigma_j^x + \hbar B \sum_i \sigma_i^z$$

where $\hbar = \hbar/2\pi$, σ_i^β ($\beta = x, y, z$) are the spin-1/2 Pauli operators for the i th spin and B is an effective transverse magnetic field strength. The coupling matrix $J_{ij} \propto 1/|i-j|^\alpha$ has an approximately power-law dependence with distance $|i-j|$, with an exponent tunable between infinite range ($\alpha = 0$) and short range ($\alpha = 3$). Previous trapped ion experiments^{23–25} have investigated the ground state properties of H_{Ising} . Here, we investigate out-of-equilibrium dynamics and a new regime where $B \gg \max(|J_{ij}|)$. In this case, the number of spin excitations (\uparrow_z) is conserved during the dynamics, and H_{Ising} reduces to the XY model of hopping hard-core bosons, $H_{\text{XY}} = \hbar \sum_{i < j} J_{ij} (\sigma_i^+ \sigma_j^- + \sigma_i^- \sigma_j^+)$. For further experiment details, see Methods.

Using a tightly focused laser and single-ion resolved imaging, we can inject localized excitations and observe the system's response. These local quenches (that is, sudden perturbations of the system) allow the dynamical properties of individual quasiparticle wave-packets to be studied, in particular their dispersion relation and entanglement distribution. Local quenches are performed by flipping one or more spins, thus coherently populating a range of quasiparticle modes. Each mode can be pictured as an equal superposition of spin waves (magnons) with positive and negative momentum $\pm k$. Therefore, in the subsequent time evolution, superpositions of left-travelling and right-travelling spin waves, ejected from the quench site, distribute entanglement amongst the spins (Fig. 1a).

The precise dynamics are determined by the quasiparticle dispersion relation, which in turn can be tuned by the spin–spin interaction range¹². We can verify the correct implementation of the desired interaction range and the corresponding quasiparticle dispersion relation by directly measuring the spatial distribution of the spin–spin interactions (Fig. 1b). The measurement closely matches theoretical predictions (see Extended Data Fig. 5). Fitting a dispersion relation from power-law interactions allows us to extract the exponent α (Fig. 1c).

The spread of information from the local quench sites can be observed in spatially and temporally resolved single-spin observables, such as the spin polarization $\langle \sigma_i^z(t) \rangle$, which characterizes the magnetization of the system (Fig. 2a–c). At early times, localized spin-wave packets radiating away from single spin-excitations are clearly visible. Later, reflections result in complex interference patterns, with properties determined by both α and the spin chain length. Questions such as whether the initially

¹Institut für Quantenoptik und Quanteninformation, Österreichische Akademie der Wissenschaften, Technikerstraße 21a, 6020 Innsbruck, Austria. ²Institut für Experimentalphysik, Universität Innsbruck, Technikerstraße 25, 6020 Innsbruck, Austria. ³Institut für Theoretische Physik, Universität Innsbruck, Technikerstraße 25, 6020 Innsbruck, Austria.

*These authors contributed equally to this work.

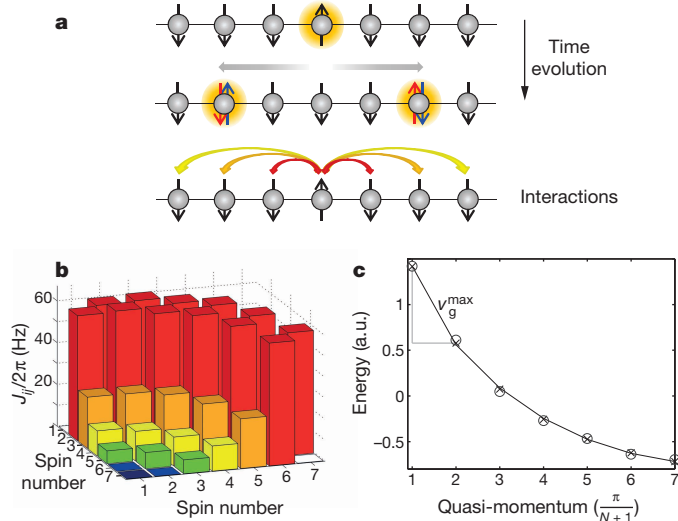


Figure 1 | Quantum dynamics in a one-dimensional spin chain following a local quench. **a**, A steady state is locally perturbed by flipping one spin. Quasiparticle wave-packets propagate left and right from the quench site and entangle spin pairs across the system. The underlying spin–spin interaction defines possible direct hopping paths (examples shown as arrows) and the quasiparticle dispersion relation. **b**, Example of a long-range spin–spin interaction matrix J_{ij} , directly measured in our system for $N = 7$ spins (see Methods), with colours matched to the interactions pictured in **a**. **c**, Quasiparticle dispersion relation (shifted by energy B), derived from **b** (circles) and predicted using experimental parameters (crosses). The line is the fitted dispersion relation for power-law interactions, with best-fit exponent $\alpha = 1.36$. The maximum group velocity v_g^{\max} is inferred from the curve’s steepest slope (we set the lattice spacing to unity). a.u., arbitrary units. Error bars (1σ) are smaller than symbols used.

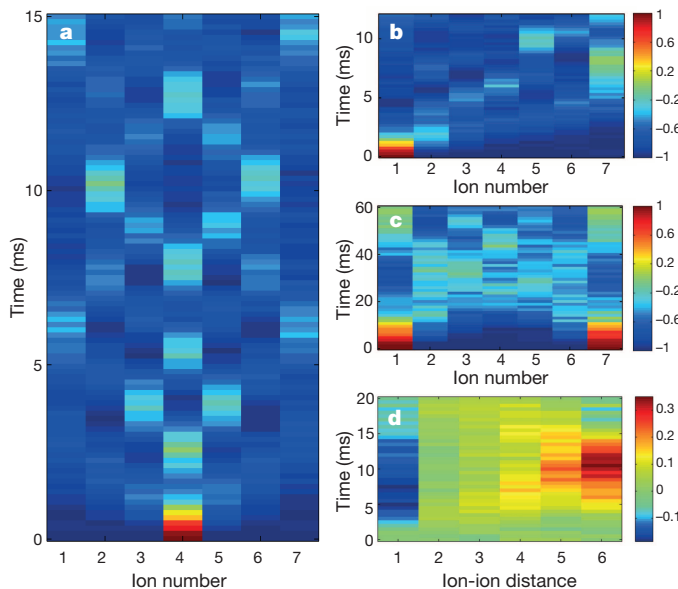


Figure 2 | Measured quantum dynamics in a seven-ion system following local and global quenches. **a–c**, Time evolution of the magnetization $\langle \sigma_i^z(t) \rangle$ (colour coded) following a local quench at: **a**, the central spin (ion), for $\alpha \approx 1.36$; **b**, the leftmost spin, for $\alpha \approx 1.36$; and **c**, both ends of the chain, for $\alpha \approx 1.75$. Values of ± 1 correspond to the fully polarized states. The colour scales in **b** and **c** also refer to **a**. **d**, Time evolution of the averaged two-spin correlation function $\bar{C}_n = \frac{1}{N-n} \sum_{i=1}^{N-n} C_{i,i+n}$ (colour coded) where $C_{ij} = \langle \sigma_i^z \sigma_j^z \rangle - \langle \sigma_i^z \rangle \langle \sigma_j^z \rangle$ following a global quench, for $\alpha \approx 1.75$. The timescales (in units of $1/J$, where $J = \sum_{i,j} J_{i,j}/(N-1)$) so that the time evolution can be compared with different values of α are for **a–d**, respectively: $T_{\max} = 8.36, 6.72, 5.44, 3.12$.

localized excitation refocusses are non-trivial even in the simpler case of nearest-neighbour interactions¹³.

Flipping several spins at both ends of the chain creates counter-propagating wavefronts, opening the prospect of studying quasiparticle collisions (Fig. 2c). Extended Data Fig. 1 shows close agreement with theory in all cases. Initializing all N spins in $|\downarrow\rangle + |\uparrow\rangle$ realizes a global quench. In this case, the many-body state is in a superposition containing 0 through to N excitations, in which interactions between single-excitation quasiparticle modes can no longer be neglected. In this case, the resulting distribution of information can be observed through two-point correlation functions^{9,26}, as seen in Fig. 2d and Extended Data Fig. 2.

To reveal the distribution of quantum correlations after a local quench, we tomographically measure the evolution of the full quantum state of pairs of spins (see Fig. 3 and Extended Data Fig. 3). Figure 3a exemplifies the results for an interaction range $\alpha \approx 1.75$, for which a clear wavefront is apparent. The results show that magnon wave-packets emerging from either side of the initial excitation distribute entanglement across the spin chain (Fig. 3b, c); the wavefront first entangles spins neighbouring the quench site, then the next-nearest neighbours, and so on until the boundaries are reached.

Finally, we investigate how the spin–spin interaction range affects the way in which information is transported around the system. For this, we measure the magnetization dynamics following a local quench in a chain of 15 spins, for three values of α roughly equally spaced around $\alpha = 1$. In the shortest-range case (Fig. 4a, $\alpha = 1.41$), an approximate light cone can be seen. There is a clear leading wavefront of spin-excitation that moves away from the quench site at a well defined velocity, and outside which the signal decays rapidly (Fig. 4a, d). These are the features of a well-defined speed limit for quantum dynamics that one would expect for finite-range interactions, and that has previously been observed in systems of neutral atoms with nearest-neighbour interactions^{9,10}. Indeed, the information transport observed in our shortest-range experiment is largely captured by a Lieb–Robinson bound that considers only the nearest-neighbour interactions in the system (Fig. 4a, d, e).

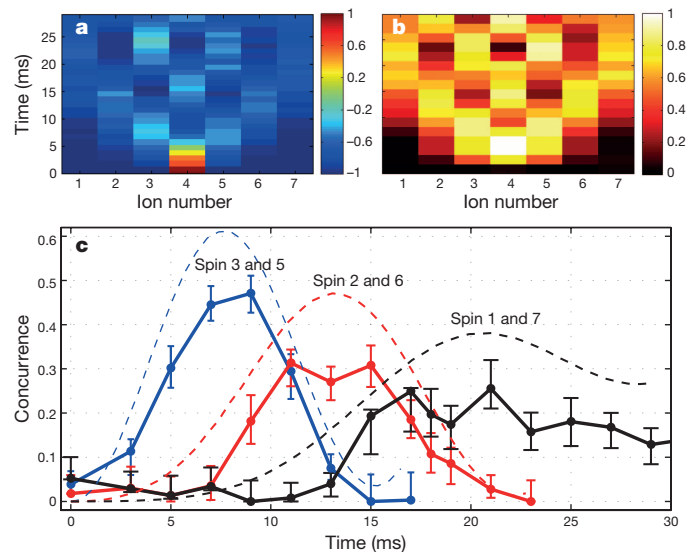


Figure 3 | Entanglement distribution following a local quench. Shown are the dynamics for $\alpha \approx 1.75$ up to $T = 30$ ms ($2.99 J^{-1}$). **a**, Measured single-spin magnetization (colour coded). **b**, Single-spin von Neumann entropy $-\text{Tr}(\rho \log(\rho))/\log(2)$ (colour coded) derived from measured density matrices. High-entropy states are due to correlations with other spins. **c**, Evolution of entanglement (concurrence, see Methods) between pairs of spins distributed symmetrically around the central spin, revealing the propagation of entangled quasiparticles from the centre to the boundaries of the system. Blue, spins 3 and 5; red, spins 2 and 6; black, spins 1 and 7. Dashed lines show theoretical predictions. Error bars, 1σ calculated via Monte Carlo simulation of quantum projection noise³⁰.

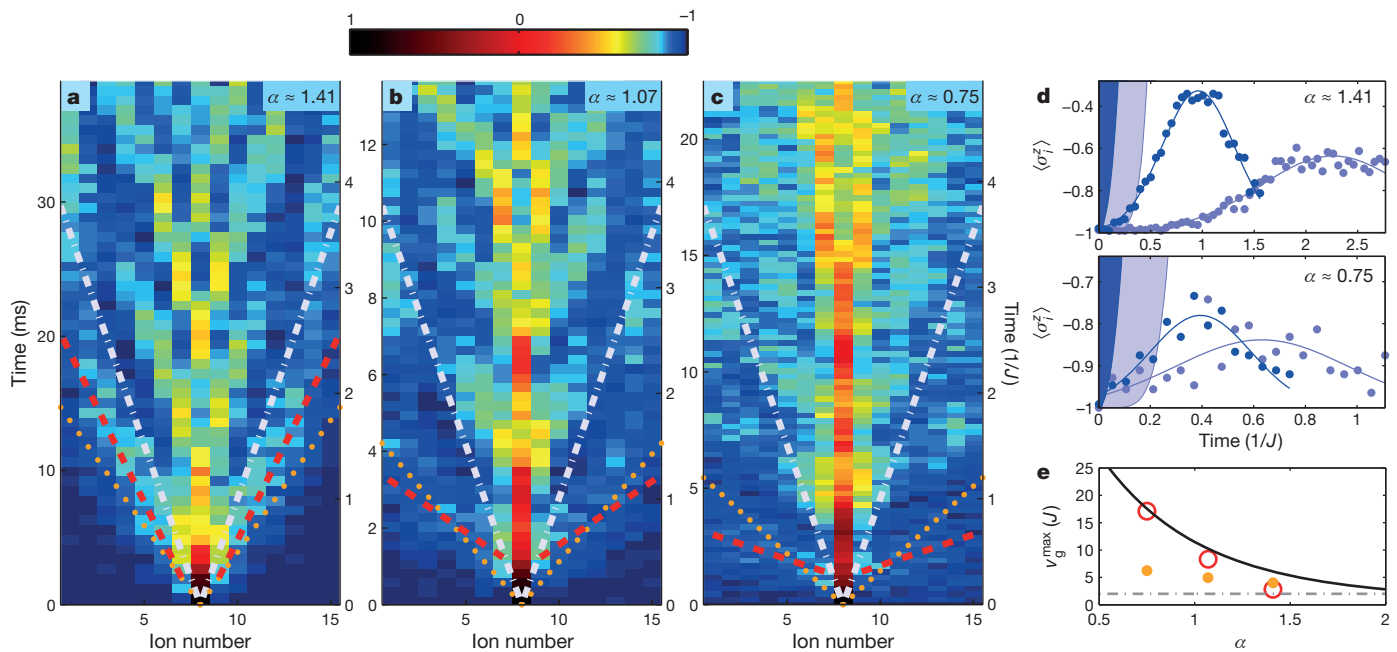


Figure 4 | Measured quantum dynamics for increasing spin-spin interaction ranges. **a–c**, Measured magnetization $\langle \sigma_i^z(t) \rangle$ (colour coded) following a local quench. From **a** to **c**, the interaction ranges are $\alpha \approx 1.41$, 1.07 , 0.75 . In **a**, an effective light cone is evident and the dynamics are approximately described by nearest-neighbour interactions only. Red lines, fits to the observed magnon arrival times (examples in **d**); white lines, light cone for averaged nearest-neighbour interactions; orange dots, after renormalization by the algebraic tail (see Methods). As the interaction range is increased (**b**, **c**) the light cone disappears and nearest-neighbour models fail to capture the dynamics. **d**, Magnetization of spins (ions) 6 and 13, from **a** (top) and **c** (bottom). Solid

As the interaction range is increased (Fig. 4b, c), the arrival times of the first maxima in magnetization are seen to appear earlier and earlier, reflecting the ejection of faster and faster quasiparticles from the quench site. Furthermore, the signal decay outside these maxima is very slow: there is an almost instant increase in the magnetization even at large distances (Fig. 4d, top). Clearly we are able to tune our system into a regime where the light-cone picture does not apply and significant amounts of information can propagate directly to distant neighbours. This is consistent with generalized Lieb–Robinson bounds for power laws, which for $\alpha < 1$ are trivial, placing no restriction on the speed of information propagation^{6–8}.

A quantitative analysis is provided by extracting the maximum quasiparticle group velocity v_g^{\max} from the data (see Methods and Extended Data Fig. 4). For the shortest-range case, the observed v_g^{\max} fits well with the nearest-neighbour case (Fig. 4d). As the interaction range is increased, the results are consistent with a divergence of v_g^{\max} , as recently predicted¹². Ultimately, the information propagation speed in our system is limited by the propagation of acoustic waves across the ion chain²¹. Note that, despite the faster-moving components in the longer-range data (Fig. 4c), the initial perturbation remains more localized. This is consistent with the predicted flattening of the dispersion relation away from the divergence. For a comparison of data with theory, see Extended Data Fig. 4.

Differences between the observed and ideal quantum dynamics following local quenches largely correspond to imperfect conservation of excitation number. This could be caused by electric field noise leading to heating of the ion’s motional state or by unwanted spin–motion entanglement. For global quench dynamics, laser-frequency and magnetic-field fluctuations give rise to dephasing.

We have presented a new platform for investigating quantum phenomena—a many-body quantum system in which the states and properties of its quasiparticle excitations can be precisely initialized, controlled and measured. This opens many new paths for experimental

investigations, the subjects of which can be broadly split into the following: (1) quantum transport phenomena, concerning how quantum states and entanglement¹³, or excitations^{14,27}, propagate across quantum many-body systems; (2) how quantum systems reach equilibrium, including the question of when thermalization^{15,28} and localization occur¹⁶; (3) entanglement growth and simulation complexity¹⁷ (the interaction range parameter α is known to play a critical role in the growth rate of entanglement and the possibility of simulating the dynamics with conventional computers); and (4) quasiparticle behaviour near phase transitions¹. For many of these research lines it would be useful, and feasible, to add localized spin excitation absorbers or reflective boundaries, and static or stochastically fluctuating disorder, to our system.

During the final stage of this work, we became aware of complementary recent work investigating global quenches of trapped-ion spin chains²⁶.

METHODS SUMMARY

Ions are held in a linear Paul trap, each encoding a spin-1/2 particle in the electronic states $|S_{1/2}, m = +1/2\rangle \equiv |\downarrow\rangle$ and $|D_{5/2}, m = +5/2\rangle \equiv |\uparrow\rangle$. Spins are manipulated with a narrow-linewidth laser at 729 nm (ref. 29). Ions are coherently manipulated with two laser beams intersecting the ion string perpendicularly from opposite directions. The first beam interacts with all the ions with nearly equal strength and is used for carrying out collective spin rotations, as well as implementing effective spin–spin interactions by means of electronic-state-dependent forces³. These forces off-resonantly drive the transverse motional modes of the ion string. The interaction range (α) is controlled by how far off-resonant the driving is and the axial trapping confinement. The second beam is strongly focused, steerable, and is used for single-spin rotations. Spatially resolved fluorescence measurements in conjunction with prior single-spin rotations allow us to take single-shot measurements of arbitrary spin correlations.

If our system had only nearest-neighbour interactions, the signal propagation after a local perturbation using σ_ℓ^z would be bounded by $|\langle \psi(t) | \mathcal{O} | \psi(t) \rangle - \langle \psi_0 | \mathcal{O} | \psi_0 \rangle| \leq 2\|\mathcal{O}\|I_d(4|t|\max_i(J_{i,i+1}))$, where \mathcal{O} may be any local operator with norm $\|\mathcal{O}\|$ and distance d to the quench site ℓ . As Fig. 4d shows for $\mathcal{O} = \sigma_i^z$, this bound is only a

good approximation for the shortest-range case measured. Additionally, for nearest-neighbour interactions, the maximal group velocity v_g^{\max} of magnon modes (that is, the largest slope of the quasiparticle dispersion relation) is finite. Figure 4a–c shows the corresponding light cone $t = d/v_g^{\max}$, outside which the signal would diminish exponentially if our system had only nearest-neighbour interactions. Clearly, longer-range interactions play an important role in our dynamics and must be taken into account.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 January; accepted 6 May 2014.

1. Sachdev, S. Quantum criticality: competing ground states in low dimensions. *Science* **288**, 475–480 (2000).
2. Calabrese, P. & Cardy, J. Time-dependence of correlation functions following a quantum quench. *Phys. Rev. Lett.* **96**, 136801 (2006).
3. Porras, D. & Cirac, J. I. Effective quantum spin systems with trapped ions. *Phys. Rev. Lett.* **92**, 207901 (2004).
4. Friedenauer, H., Schmitz, H., Glueckert, J., Porras, D. & Schaetz, T. Simulating a quantum magnet with trapped ions. *Nature Phys.* **4**, 757–761 (2008).
5. Lieb, E. & Robinson, D. The finite group velocity of quantum spin systems. *Commun. Math. Phys.* **28**, 251–257 (1972).
6. Cramer, M., Serafini, A. & Eisert, J. in *Quantum Information and Many Body Quantum Systems* (eds Ericsson, M. & Montangero, S.) 51 (Edizioni della Normale, 2008).
7. Hastings, M. B. in *Quantum Theory from Small to Large Scales, Lecture Notes of the Les Houches Summer School* (eds Frohlich, J., Salmhofer, M., Mastropietro, V., Roock, W. D. & Cugliandolo, L.) Vol. 95, Ch. 5 (Oxford Univ. Press, 2010).
8. Nachtergaele, B. & Sims, R. Much ado about something: why Lieb-Robinson bounds are useful. *IAMP News Bull.* 22–29 (October 2010).
9. Cheneau, M. *et al.* Light-cone-like spreading of correlations in a quantum many-body system. *Nature* **481**, 484–487 (2012).
10. Fukuhara, T. *et al.* Microscopic observation of magnon bound states and their dynamics. *Nature* **502**, 76–79 (2013).
11. Langen, T., Geiger, R., Kuhnert, M., Rauer, B. & Schmiedmayer, J. Local emergence of thermal correlations in an isolated quantum many-body system. *Nature Phys.* **9**, 640–643 (2013).
12. Hauke, P. & Tagliacozzo, L. Spread of correlations in long-range interacting quantum systems. *Phys. Rev. Lett.* **111**, 207202 (2013).
13. Bose, S. Quantum communication through spin chain dynamics: an introductory overview. *Contemp. Phys.* **48**, 13–30 (2007).
14. Rebertus, P., Mohseni, M., Kassal, I., Lloyd, S. & Aspuru-Guzik, A. Environment-assisted quantum transport. *New J. Phys.* **11**, 033003 (2009).
15. Rigol, M., Dunjko, V., Yurovsky, V. & Olshanii, M. Relaxation in a completely integrable many-body quantum system: an ab initio study of the dynamics of the highly excited states of 1D lattice hard-core bosons. *Phys. Rev. Lett.* **98**, 050405 (2007).
16. Yao, N. Y. *et al.* Many-body localization with dipoles. Preprint at <http://arXiv.org/abs/1311.7151> (2014).
17. Schachenmayer, J., Lanyon, B. P., Roos, C. F. & Daley, A. J. Entanglement growth in quench dynamics with variable range interactions. *Phys. Rev. X* **3**, 031015 (2013).
18. Kitaev, A. Fault-tolerant quantum computation by anyons. *Ann. Phys.* **303**, 2–30 (2003).
19. Bravyi, S., Hastings, M. B. & Verstraete, F. Lieb-Robinson bounds and the generation of correlations and topological quantum order. *Phys. Rev. Lett.* **97**, 050401 (2006).
20. Eisert, J. & Osborne, T. General entanglement scaling laws from time evolution. *Phys. Rev. Lett.* **97**, 150404 (2006).
21. Jünemann, J., Cadarso, A., Perez-Garcia, D., Bermudez, A. & Garcia-Ripoll, J. Lieb-Robinson bounds for spin-boson lattice models and trapped ions. *Phys. Rev. Lett.* **111**, 230404 (2013).
22. Eisert, J., van den Worm, M., Manmana, S. R. & Kastner, M. Breakdown of quasi-locality in long-range quantum lattice models. *Phys. Rev. Lett.* **111**, 260401 (2013).
23. Kim, K. *et al.* Entanglement and tunable spin-spin couplings between trapped ions using multiple transverse modes. *Phys. Rev. Lett.* **103**, 120502 (2009).
24. Britton, J. W. *et al.* Engineered two-dimensional Ising interactions in a trapped-ion quantum simulator with hundreds of spins. *Nature* **484**, 489–492 (2012).
25. Islam, R. *et al.* Emergence and frustration of magnetism with variable-range interactions in a quantum simulator. *Science* **340**, 583–587 (2013).
26. Richerme, P. *et al.* Non-local propagation of correlations in long-range interacting quantum systems. *Nature* <http://dx.doi.org/10.1038/nature13450> (this issue).
27. Plenio, M. B. & Huelga, S. F. Dephasing-assisted transport: quantum networks and biomolecules. *New J. Phys.* **10**, 113019 (2008).
28. Gong, Z.-X. & Duan, L.-M. Prethermalization and dynamical transition in an isolated trapped ion spin chain. *New J. Phys.* **15**, 113051 (2013).
29. Schindler, P. *et al.* A quantum information processor with trapped ions. *New J. Phys.* **15**, 123012 (2013).
30. Roos, C. F. *et al.* Bell states of atoms with ultralong lifetimes and their tomographic state analysis. *Phys. Rev. Lett.* **92**, 220402 (2004).

Acknowledgements We acknowledge discussions with L. Tagliacozzo, M. Heyl, A. Gorshkov and S. Bose. This work was supported by the Austrian Science Fund (FWF) under grant number P25354-N20, and by the European Commission via the integrated project SIQS and by the Institut für Quanteninformation. We also acknowledge support from the European Research Council through the CRYTERION Project (number 227959).

Author Contributions P.H., B.P.L. and C.F.R. developed the research, based on theoretical ideas conceived with P.Z.; P.J., B.P.L., C. H. and C.F.R. performed the experiments; B.P.L., P.J., C.F.R. and P.H. analysed the data and carried out numerical simulations. P.J., C.H., B.P.L., R.B. and C.F.R. contributed to the experiment; B.P.L., C.F.R., P.H., P.Z. and R.B. wrote the manuscript; all authors contributed to discussions of the results and of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.F.R. (christian.roos@uibk.ac.at).

An optoelectronic framework enabled by low-dimensional phase-change films

Peiman Hosseini¹, C. David Wright² & Harish Bhaskaran¹

The development of materials whose refractive index can be optically transformed as desired, such as chalcogenide-based phase-change materials, has revolutionized the media and data storage industries by providing inexpensive, high-speed, portable and reliable platforms able to store vast quantities of data. Phase-change materials switch between two solid states—amorphous and crystalline—in response to a stimulus, such as heat, with an associated change in the physical properties of the material, including optical absorption, electrical conductance and Young's modulus^{1–5}. The initial applications of these materials (particularly the germanium antimony tellurium alloy Ge₂Sb₂Te₅) exploited the reversible change in their optical properties in rewritable optical data storage technologies^{6,7}. More recently, the change in their electrical conductivity has also been extensively studied in the development of non-volatile phase-change memories^{4,5}. Here we show that by combining the optical and electronic property modulation of such materials, display and data visualization applications that go beyond data storage can be created. Using extremely thin phase-change materials and transparent conductors, we demonstrate electrically induced stable colour changes in both reflective and semi-transparent modes. Further, we show how a pixelated approach can be used in displays on both rigid and flexible films. This optoelectronic framework using low-dimensional phase-change materials has many likely applications, such as ultrafast, entirely solid-state displays with nanometre-scale pixels, semi-transparent 'smart' glasses, 'smart' contact lenses and artificial retina devices.

Phase-change materials (PCMs) have been key to enabling the creation of a versatile platform able to store vast amounts of information on inexpensive plastic substrates (DVDs and other optical media^{6,7}) using optical pulses. These technologies use light of specific wavelengths (for example, Blu-Ray at 405 nm, DVDs at 650 nm and CDs at 780 nm) to write and read data. The next generation of electronic, solid-state memories based on PCMs could replace the current leading storage technologies, namely Flash and magnetic disks. PCMs^{1–5} have extraordinary properties such as extreme scalability, fast switching speeds, and high switching endurance^{4,5}. However, although PCMs have been commercialized in both the electrical⁴ and optical domains⁸, there has been very little research into understanding how transforming the phase in one domain affects the properties in another (that is, how switching electrically affects optical properties, and vice versa). This is because the likely applications emerging from such non-volatile, optoelectronic properties have so far been unclear. Here we propose and demonstrate these optoelectronic properties in the context of novel display and data visualization applications.

Recently, thin-film perfect absorbers in the mid-infrared part of the spectrum have been demonstrated, where the optical properties of a temperature-transient PCM (vanadium dioxide, VO₂) were tuned⁹. VO₂, however, is stable in each of its phases only within certain temperature ranges, a property that makes it energy inefficient and impractical to implement in electrically controlled circuits at the nanoscale. In contrast, the primary phases of Ge₂Sb₂Te₅ (GST)—amorphous and crystalline—are stable for most applications¹⁰ and can be thermally, optically or electrically switched at ultrahigh speed^{11,12}. Furthermore, GST is extremely

scalable^{13,14}, easily integrated in commercial devices and has optical properties that can be tuned in the visible spectrum of light¹⁵.

We describe a unique optoelectronic framework technology employing GST-based thin films and experimentally demonstrate its wide-ranging applicability in various types of display technologies. We show how such a system, when combined with a transparent electrode such as indium tin oxide (ITO), can be used as a reflective or a semi-transparent microdisplay and on both rigid and flexible substrates, which could usher in an era of truly flexible electronic screens that can operate at the maximum possible resolution allowed by the diffraction of visible light, both as backlit displays and electronic readers. Current microdisplay technologies based on liquid crystals, microelectromechanical systems and organic light-emitting diodes are attracting considerable attention because of a growing interest in wearable technology¹⁶. Key requirements for such applications are high resolution, high speed and low power consumption, all of which are met by the technology described here.

We first describe the reflective type of display using GST-based thin films. Figure 1a shows the cross-sectional schematic of our reflective devices; the GST is sandwiched between two ITO layers and deposited on top of a reflective surface such as platinum. The crystallization of the few-nanometres-thick GST layer induces a colour change in the entire film, visible when incident 'white' light is reflected back. To understand the relationship between the thickness of the ITO and GST layers and the overall optical properties of the stack we systematically compute the reflectivity spectra of the stack while gradually increasing the thickness of each layer. Every calculation is made twice, one for each phase of the GST layer; a change in the refractive index of the GST layer (caused by a change in phase) modulates the reflective spectrum of the entire stack. The spectra are finally compared and the percentage change in reflectivity is plotted

as $\Delta R (\%) = \left(\frac{R_{\text{crystal}} - R_{\text{amorphous}}}{R_{\text{amorphous}}} \times 100 \right)$, with R_{crystal} being the reflectivity (at a specific wavelength) of the entire stack when the GST layer is in the crystalline phase and $R_{\text{amorphous}}$ its reflectivity when the GST layer is in the amorphous phase. We use the above expression because the amorphous phase is the as-deposited phase in our experiments. These models (top graphs in Fig. 1b) indicate that one key factor is the presence of a transparent conductor beneath the GST (with thickness t in Fig. 1a), which we refer to as the ITO spacer. This layer is crucial and contributes to the colour in our system.

To experimentally verify our reflectivity calculations, we sputter a number of multi-layered films on silicon wafers with a 300-nm-thick thermally grown silicon dioxide (SiO₂) layer; our GST layer is known to be amorphous as sputtered (section S4 in the Supplementary Information). Figure 1d shows a photograph of a few sputtered films with similar structure (10 nm ITO/7 nm amorphous GST/ t nm ITO/100 nm Pt/SiO₂); t is the only thickness that is varied. The colour of the films when t (the ITO spacer thickness) is varied varies dramatically. To fully crystallize the GST, we heat each sample up to 220 °C for a few minutes on a hot plate. The observed variation in the colour of these samples when the GST is crystallized is also shown in Fig. 1d; the colour variations show the

¹Department of Materials, University of Oxford, Parks Road, Oxford OX1 3PH, UK. ²College of Engineering, Mathematics and Physical Sciences, University of Exeter, Harrison Building, North Park Road, Exeter EX4 4QF, UK.

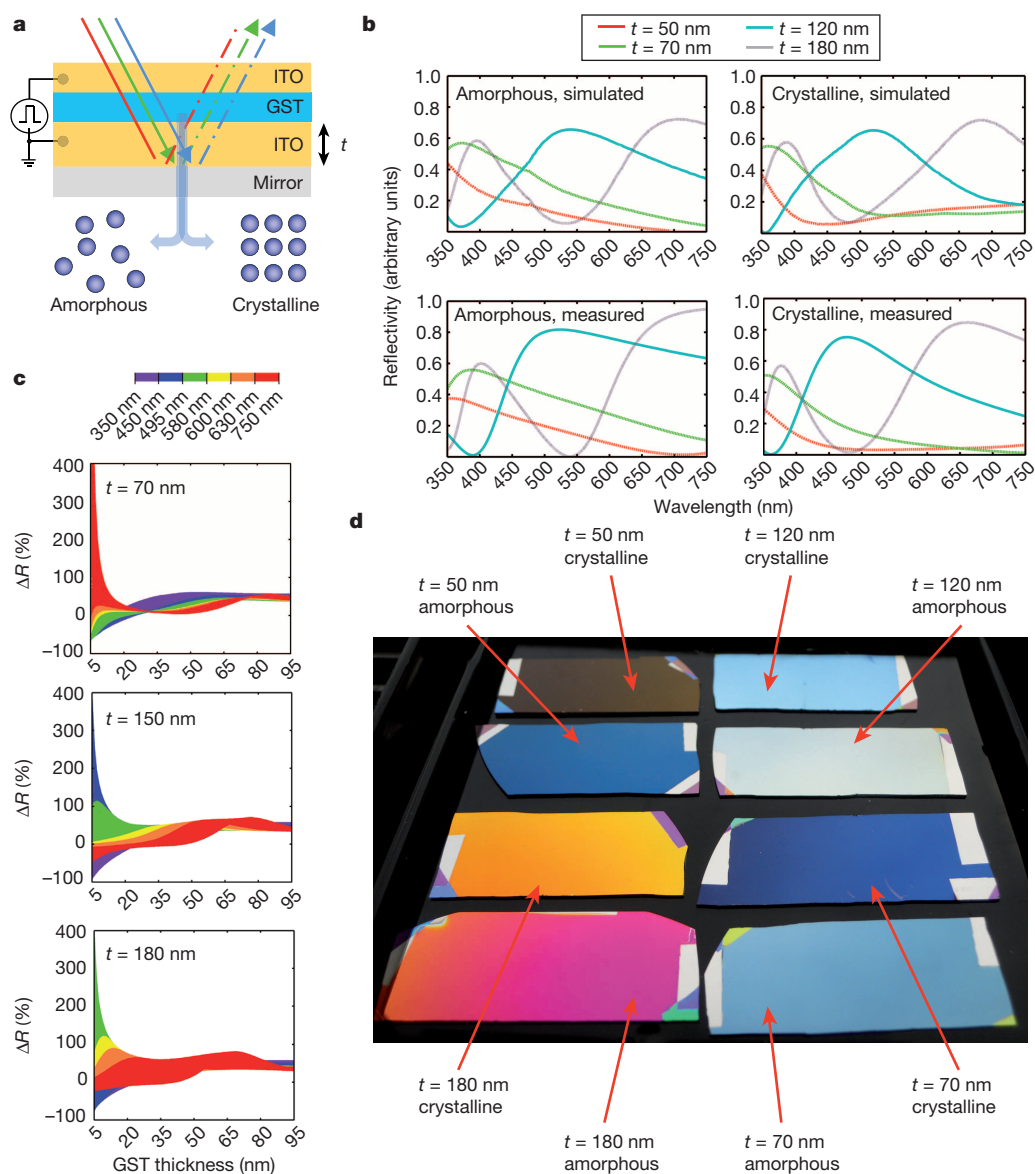


Figure 1 | Colour tunability using ultrathin PCM films. **a**, Schematic representation of the thin-film material stack comprising ITO/GST/ITO. The crystallization of the few-nanometres-thick GST layer induces a colour change in the entire film, visible when incident white light is reflected back. **b**, Simulated (top, left) and measured (bottom left) reflectivity spectra of the amorphous samples together with the simulated (top, right) and measured (bottom, right) reflectivity spectra of the crystalline samples shown in **d**. In both phases, the close match between the model and simulation is striking. **c**, Change in reflectivity simulated for visible wavelengths, for various thicknesses of the ITO spacer and with increasing thickness of the GST layers. Each graph represents the change in optical reflectivity $(R_{\text{crystal}} - R_{\text{amorphous}})/R_{\text{amorphous}} \times 100$ when the PCM embedded in the ITO/GST/ITO/Pt structure is switched from the amorphous to the crystalline phase. By tailoring the

thickness of the ITO spacer, the reflectivity of specific wavelengths is enhanced, and the remaining wavelengths are either decreased or weakly increased, when white light is used for illumination. **d**, An example of four different films with similar layered structure (starting from the topmost layer) consisting of 10 nm ITO/7 nm GST/ t nm ITO/100 nm Pt with different thicknesses of t , the ITO layer deposited between the GST and the reflective Pt layer. The value of t determines the reflected colour. After the deposition process, one part of each sample is placed on a hot plate at 220 °C for a few minutes to induce complete crystallization of the GST layer. The corresponding colour change is also visible, and is quite dramatic. For example, the white ($t = 120$ nm) amorphous film turns light blue when crystallized. No postproduction colour is added to enhance contrast.

potential for the use of such films as individual pixels in display devices, each pixel being a different colour to achieve a standard colour display device (and with switching in real devices achieved electrically; see below).

To quantify this colour change, we first model the system using a transfer matrix optical computational method¹⁷ to calculate properties such as reflectance, transmittance and internal electrical field at visible wavelengths (full details of the model are available in section S5 of the Supplementary Information). We then carry out reflectivity measurements on sputtered films to compare our model with experiments; we use a spectrometer (Lambda 1050, Perkin Elmer US) to obtain the reflectivity spectra of the samples both before and after crystallization. Figure 1b

shows exceptionally good agreement between the simulated (top, left) and measured (bottom, left) reflectivity spectra for the samples shown in Fig. 1d, when the thin GST layer is in the amorphous phase. Similar reflectivity spectra are shown (Fig. 1b, right) for the same samples after crystallizing the GST layer, where again the agreement between experiment and theory is very good. These results show that an effective change in the reflectivity spectrum (that is, colour) of the stack can be induced simply by changing the phase of a chalcogenide layer a few nanometres thick.

The next aspect of our study focuses on the influence of the GST thickness. In conventional optical disks, the thickness of the GST layer is

optimized, along with the thicknesses of surrounding dielectric layers, to enhance the optical contrast. However, the GST thicknesses used in commercial disks are relatively high (several tens of nanometres, typically). In the case of the films shown in Fig. 1d, the thickness of the GST layer is just 7 nm (the very minimum we could sputter reliably using our facilities). Counter-intuitively, as we explain below, smaller thicknesses enhance contrast dramatically and, importantly, lead to lower power electrical switching. Figure 1c shows our computational study of the change (shown as a percentage) in optical reflectivity for various thicknesses of the ITO spacer and with increasing thickness of the phase-change layer. For a thin layer of GST (that is, below 20 nm) the range of wavelengths that show increased reflectivity upon crystallization varies depending on the chosen thickness of the ITO spacer. This allows the perceived colour of the entire stack to be transformed simply by changing the phase of a GST layer a few nanometres thick. Interestingly, this crystallization-induced variation of reflectivity is exponentially amplified for thinner layers of the PCM. Also, by tailoring the thickness of the ITO spacer, the reflectivity of specific wavelengths is enhanced when white light is used as the illumination source. Figure 1c shows three examples with ITO thicknesses of 70 nm, 150 nm and 180 nm where the red, green and blue components, respectively, are enhanced with the remaining wavelengths either decreased or weakly increased. The importance of the ITO spacer is further elucidated in section S7 of the Supplementary Information, where the absence of interference (that is, no ITO spacer is inserted between the GST and the Pt layer) causes the reflectivity of the film to increase only weakly in the crystalline phase, mostly independently of the wavelength.

We now demonstrate how this thin-film stack could be used in future PCM-based display applications. To demonstrate the working principle of a display that can be electrically switched, we pattern an image using a

nanometre-sized electrical probe (that is, the conductive tip of an atomic force microscope, CAFM), which was previously shown to be capable of switching nanoscale crystalline areas^{18–20} in PCM films. Our technique uses a pixelated approach mimicking real pixels in actual devices (the fabrication of such a device is described below). The use of CAFM to characterize the nanoscale switching characteristics of PCM from the more-amorphous to more-crystalline phase has been studied previously, and is known to closely resemble a standard cross-bar phase-change memory cell; hence, we employ this technique to demonstrate our electrically enabled display technology^{18,21}. More details regarding this technique and how it relates to actual PCM-based devices (pixels) can be found in section S1 of the Supplementary Information. To demonstrate the high optical contrast, we start by electrically drawing the Oxford Radcliffe Camera building (the grayscale image of which is shown in Fig. 2b) in Fig. 2c, d and e on similar films with ITO spacer thicknesses of 50 nm, 70 nm and 180 nm, respectively. All the pictures are clearly visible using a standard, non-polarized, optical microscope. No artificial contrast or colour has been added in postproduction. The striking contrast demonstrates the feasibility of such ultrathin PCM-based microdisplays. To demonstrate the colour-rendering capability, a different pattern representing the University of Oxford logo (shown in Fig. 2f) is also electrically drawn on a 20 nm ITO/7 nm GST/70 nm ITO/100 nm Pt stack. The resulting optical image is shown in Fig. 2g, which closely matches the original logo, representing the ‘Oxford blue’ in Fig. 2f (note the striking contrast between the blue and white regions).

We then verify that a pixelated array, as would be necessary in a display where each pixel can be randomly accessed and manipulated, is approximated very closely by our CAFM approach. To perform this verification, we fabricate an array of 300 nm × 300 nm pixels with 200 nm pitch on a

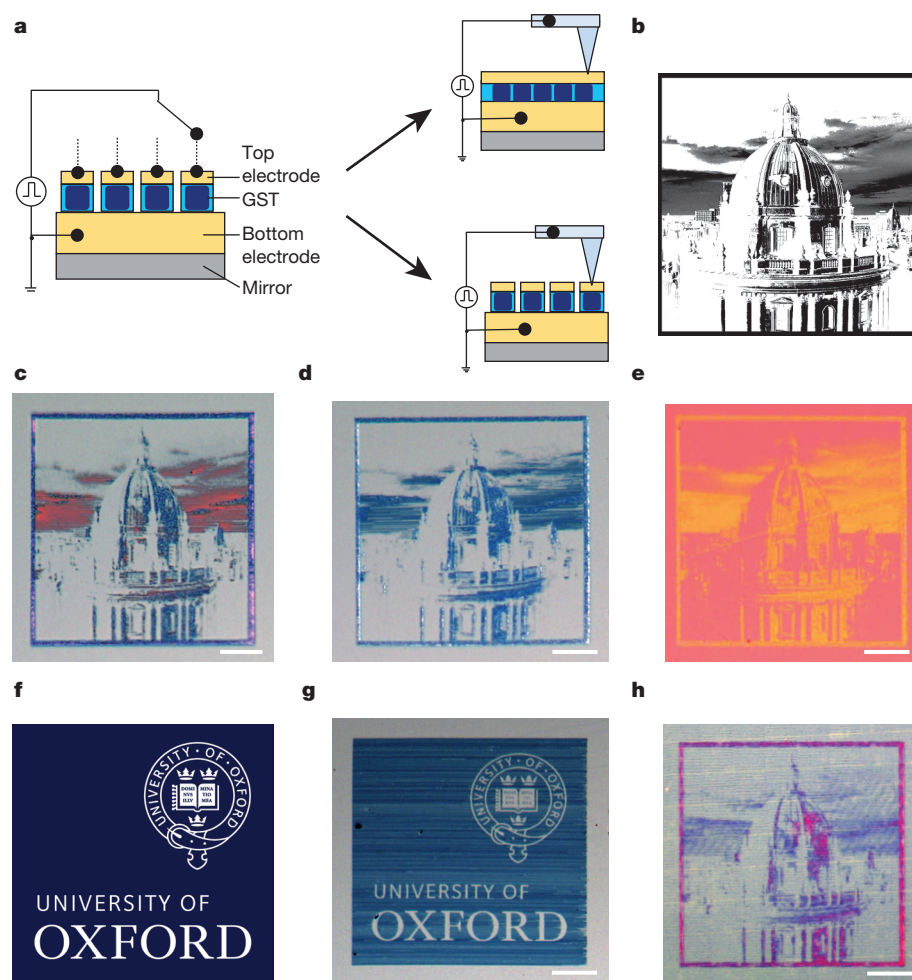


Figure 2 | Reflective display films. **a**, Schematic illustration of the electrically induced colour changes in phase-change-based electronic displays. **b**, The original (grayscale) picture of the Oxford Radcliffe Camera used as a pattern. **c**, Electrically constructed image on a continuous ITO/GST/ITO/Pt stack with a 50-nm-thick ITO spacer; the white regions are regions where the phase is amorphous. No artificial contrast has been added in postproduction. The pattern is once again electrically reproduced on various stacks with ITO spacer thickness of 70 nm (**d**) and 180 nm (**e**). Scale bars are 10 μ m, which show the ultimate resolution such displays are capable of achieving. **f**, The original image of the University of Oxford logo is used as a different pattern. **g**, Electrically constructed image on a ITO/GST/ITO/Pt stack with 70-nm-thick ITO spacer—the blue regions are regions where the phase is transformed. **h**, To verify the effectiveness of our blanket film approach we electrically transformed an array of lithographically defined 300 nm × 300 nm pixels (200 nm pitch) on a standard vertical stack with a 50-nm-thick ITO spacer; the optical contrast is strikingly evident, experimentally confirming that the pixelated array is well approximated by a CAFM-generated one.

standard vertical stack with a 50-nm-thick ITO spacer. The active volume of each phase-change cell (that is, the GST layer) is restricted to a single pixel precisely addressable by the nanoscale tip. These pixels were electrically switched using a nanoscale conductive tip. Figure 2h shows the resulting optical image, where the optical contrast is once again striking. Further experiments using electrically switched single pixels, together with the resulting high optical contrast, are shown in section S1 of the Supplementary Information.

We now move on to transparent substrates and electrically modulate their optical transmission. Future semi-transparent displays will find applications in emerging technologies such as contact-lens-type displays²², 'smart' eye glasses, windshield displays, for modulating light transmission through windows²³ or even for synthetic retina devices²⁴. Such light modulation, achievable with resolution beyond the diffraction limit, is also important for future ultrafast phase modulators, where a white light behind the semi-transparent display is selectively diffracted to produce interference. For such a configuration the reflective platinum underlayer is not deposited and the entire structure (ITO/GST/ITO) is sputtered directly on top of a transparent substrate, quartz in our case. As in the previous case of reflective displays, the electrically induced crystallization (and re-amorphization) can be used as a means of creating optical contrast.

Figure 3a shows a schematic of the principle behind semi-transparent, phase-change-based displays. The calculated change in transmission (shown as a percentage) between the amorphous and crystalline phase is shown in Fig. 3b. According to the model, a GST thickness of around 75 nm to 85 nm gives the highest change in transmission for most wavelengths. However, owing to the intrinsic absorption of the GST layer, the

opacity of the films quickly drops after a few nanometres (see section S6 of the Supplementary Information). We therefore choose a 7 nm GST layer sandwiched between a 20-nm-thick ITO and a 40-nm-thick ITO as the best compromise between contrast and transparency. We also confirm the calculated change in absorption and transmission characteristics of such a film (shown in Fig. 3c) by measuring and comparing its transmission (or absorption) spectra upon crystallization.

We use the CAFM (similar to the experiments on reflective substrates) to render pixelated images on these substrates. Figure 3e shows an example of an electrically constructed image on such a substrate. To illustrate the transmissivity of the substrate, a square green pad is placed underneath the rendered image; the transmissivity modulation is visible in Fig. 3d and 3e (the logo of Wolfson College, University of Oxford) and the contrast and transparency of the electrically drawn image is clear. This demonstrates the principle of using such phase-change optoelectronic frameworks for transmissive micro- and nanodisplay applications, from electronic displays on windows, to backlit displays and as future wavelength-tunable windows.

Finally, given the extremely low thickness of our films, we find that they are amenable to flexible films (and therefore flexible displays). The development of a low-power, ultrathin flexible display has been a long-term goal worldwide, and the use of ITO could enhance the possibility of using them as touch-sensitive displays. Figure 4a shows a picture of a 20 nm ITO/7 nm GST/50 nm ITO/100 nm Pt multi-layered film on biaxially oriented polyethylene terephthalate (boPET). The flexibility of these films with the sputtered layers is demonstrated in Fig. 4a, c and d. The colour variation in these flexible reflective displays is entirely similar to that on rigid substrates (Fig. 2). Again, we render images using a

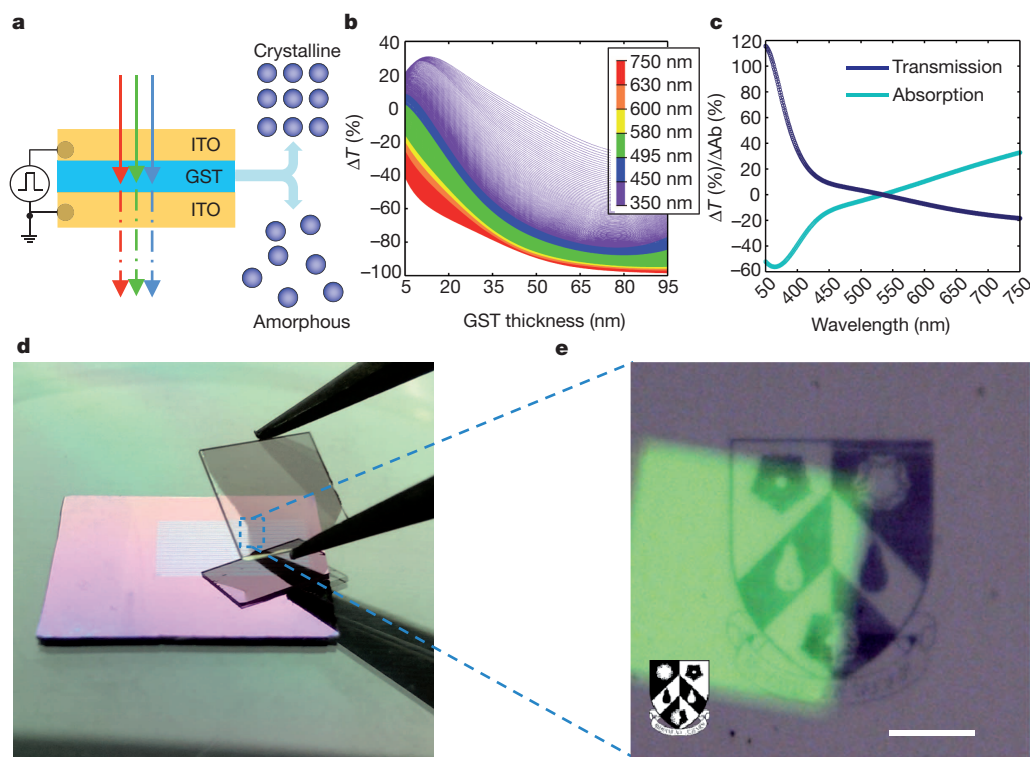


Figure 3 | Semi-transparent display type films. **a**, Schematic diagram of the principle behind a semi-transparent device using a ITO/GST/ITO structure on quartz. **b**, Computed change in optical transmission

$\Delta T(\%) = \left(\frac{T_{\text{crystal}} - T_{\text{amorphous}}}{T_{\text{amorphous}}} \times 100 \right)$, at different wavelengths, when the GST layer is switched between the amorphous and crystalline phases.

c, Measured change, upon crystallization, in absorption

$\Delta Ab(\%) = \left(\frac{Ab_{\text{crystal}} - Ab_{\text{amorphous}}}{Ab_{\text{amorphous}}} \times 100 \right)$ and transmission spectra of a

20 nm ITO/7 nm GST/40 nm ITO film sputtered on quartz. **d**, A picture of the film measured in **c**. **e**, Microscope picture of an electrically constructed image on the film, demonstrating the principle of semi-transparent PCM-based electronic paper displays. The green sample pad visible underneath the image demonstrates the semi-transparent nature of the thin-film layers and enhances clarity (because of the finite thickness of our transparent substrate, it is not in focus). The scale bar is 50 μ m, highlighting the high resolution that is possible.

nanoscale conductive probe, and the resulting image is shown in Fig. 4b. Although the quality of the boPET substrate is much inferior to that of semiconductor-grade SiO₂ or quartz, the electrically written high-resolution image of the Oxford Radcliffe Camera is readily visible. Another example of a similar film with a 180-nm-thick ITO spacer is shown in Fig. 4c, where the colour is now closer to yellow, which shows the range of possible colours. Thus the use of such flexible films in semi-transparent modes is also possible. It is also remarkable that there is no visible colour variation when the films are bent, suggesting that such thin films also can be used as wide-viewing-angle displays. A semi-transparent, flexible film is shown in Fig. 4d, completing the demonstration for reflective and semi-transparent type displays on rigid, flexible and transparent substrates.

In all the above examples we use a CAFM to electrically switch the GST sandwiched between two ITO layers. Because no device using ITO electrodes has yet been demonstrated, we show that such a device (a single pixel) can be made and switched. Using a combination of electron beam lithography, reactive ion etching and sputtering deposition, we fabricate vertical, crossbar-like ITO/GST/ITO devices to test their electrical switching characteristics. We fabricate these devices on 300-nm-thick SiO₂ on silicon substrates (see section S3 of the Supplementary Information for fabrication details, and section S1 for the use of ITO as active electrode). An optical image of a few crossbar devices is shown in Fig. 5a with a false-colour scanning electron microscopy image of one such device in the inset. This particular device has an active area (defined by the area of overlapping crossbars) of 300 nm × 300 nm. We note that switching in phase-change memories is often restricted to nanometric regions of the cell²⁵. The use of nanoscale devices (pixels) is therefore preferable, since full crystallization and re-amorphization of the pixel is required to enhance these novel optoelectronic functionalities. The electrical switching characteristics of such devices have been tested for various test cells with an example shown in Fig. 5b and c. This particular cell shows a threshold voltage²⁶ of 2.2 V with a 350× increase in conductance between the amorphous and crystalline phases. Figure 5b shows a collection of direct-current (d.c.) current–voltage curves used to transform the device to the high-conductance (low resistance or crystalline phase) state (referred to in PCM devices as the SET state). 100-ns

pulses with a fall-time of 5 ns and amplitude of 5 V are used to transform the device to its initial low conductance (high resistance or amorphous) state (referred to as the RESET state in PCM devices). A collection of SET/RESET iterations is shown in Fig. 5c demonstrating its repeatability (the cycle endurance of our devices will require further engineering to create a commercial system, but we note that commercial-type phase-change memories with endurences of greater than 10⁸ cycles have been demonstrated²⁷). It is clear that our device is a single pixel (of very high resolution), and it is reasonable to expect that such a pixel can be replicated for display applications.

We now address the energy requirements for such a display using the results obtained from the single pixel device. In phase-change memories the highest energy consumption is in the re-amorphization of a cell where sufficient energy is required to melt the PCM⁴. We estimate the energy required to re-amorphize a single ITO/GST/ITO cell to be 15 pJ, on the basis of our experimental results. A 300-cm² display with 200-nm cell size and a 50-nm gap between each cell would thus require 7.2 J to completely re-amorphize, similar to electrochromic devices²⁸. The development of even thinner PCM would enable the energy consumption of a device (pixel) to be drastically reduced by ten to a hundred times¹³, which would also increase the optical contrast, as already discussed. This would enable the use of phase-change-based electronic paper also as dynamic displays with few constraints on power requirements and ultrahigh (in the megahertz range) switching speed. Crucially for power consumption, no energy is required after a switching event, giving net zero-power consumption in static mode, a further improvement over existing colour back-lit microdisplays.

Such cells can thus be fabricated, are reliable and can be switched reversibly with low power. Many challenges remain, especially in understanding how larger pixels can be switched most effectively and yet display optical contrast, as well as in arraying and subdividing pixels effectively to create any arbitrary colour and greyscale. The use of growth-dominated PCMs may prove to be a very useful approach in this respect, although

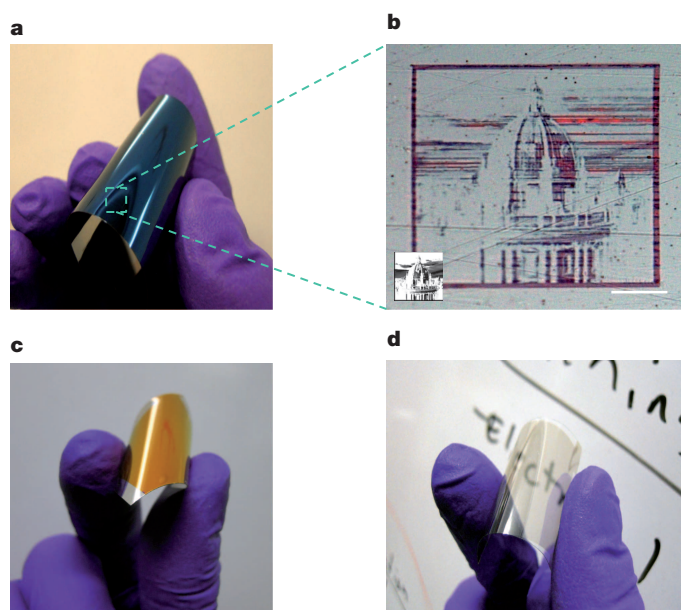


Figure 4 | Flexible display films in both reflective and semi-transparent mode. **a**, Reflective device on boPET, showing flexibility as well as the wide-viewing-angle colour consistency of such films (that is, when the film is bent, the colour in the bent region is not dramatically different). **b**, An electrically defined image is patterned directly on top of the flexible substrate using a CAFM. **c**, Another example of a reflective stack. **d**, A semi-transparent type structure sputtered directly on a flexible boPET substrate.

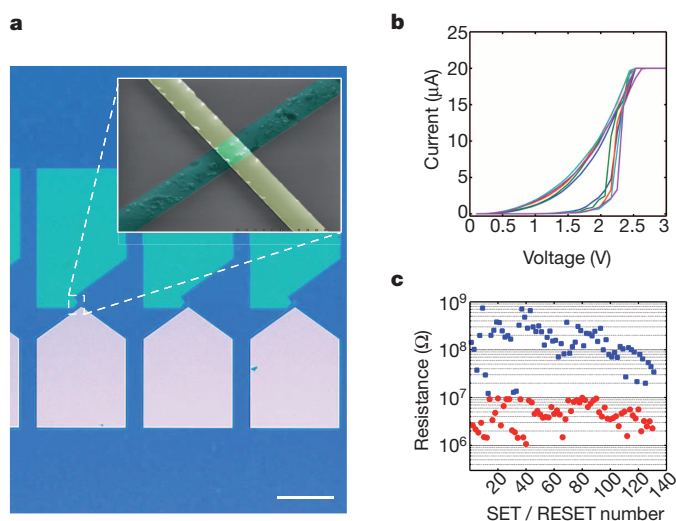


Figure 5 | Demonstration of a single pixel. **a**, Optical image of a few ITO/GST/ITO crossbar-like devices, with a false-colour scanning electron micrograph of one single device in the inset. The scale bar is 100 μm. **b**, A collection of various current–voltage curves demonstrating the amorphous-to-crystalline switching behaviour of several ITO/GST/ITO cells. These devices display a threshold voltage of approximately 2.2 V with a decrease in resistance of about 350× when switched from the amorphous to the crystalline phase. The maximum current was limited to 20 μA. **c**, SET/RESET cycling of the same device. Direct-current sweeps (similar to those in **b** above) are used to switch the device from the RESET (high resistance) state to the SET (low resistance) state. Several pulses of 5 V in amplitude with 100 ns width, 20 ns rising and 5 ns trailing edges are used to switch the device back from the SET state to the RESET state.

their optoelectronic properties also remain unexplored and could be the subject of further research in this field.

In conclusion, we have demonstrated a unique optoelectronic framework based on low-dimensional PCM films. The integration of a PCM as an active switching layer in displays paves the way for a new class of 'smart' devices that are both electrically and optically active. Our results represent the first applications of the electro-optical properties of such materials. Very thin layers of PCM can display very high optical contrast upon switching between the amorphous and crystalline phases when layered in appropriately designed thin film structures. Because this can be done electrically, it allows for the fabrication of large arrays to create displays with ultrahigh resolution. We demonstrate such displays on reflective and transparent substrates both on rigid and flexible surfaces; results that promise truly foldable ultralight displays that consume no power in static mode. We then further demonstrate the fabrication of one single pixel of a phase-change cell with transparent electrodes. Because PCMs are also capable of storing data and performing arithmetic and logic, such devices could potentially mimic the functionality of photoreceptor cells in the human eye, opening up new fields of research in human-machine interaction, multi-functional glasses, contact lenses and synthetic retina devices.

METHODS SUMMARY

Film deposition. Films are sputtered directly on thermally grown SiO₂ wafers (IDB Technology, UK), boPET (Mylar) or quartz. Substrates were cleaned for 10 min in acetone under ultrasonic agitation, rinsed in isopropanol and dried with pressurized nitrogen. Where a reflector was needed (reflective devices), 100-nm-thick Pt was deposited using a Nordiko sputtering system: starting pressure 4.7×10^{-7} torr, working pressure 1.6 mtorr, 30 sccm (standard cubic centimetres per minute) Ar, 400 W d.c. 9 r.p.m. The samples were then moved to a second sputtering system for ITO and GST deposition. ITO is sputtered from a solid target (Testbourne, UK) using 25 W direct current, 1.1 mtorr working pressure, 4×10^{-6} torr starting pressure, 100 °C sample temperature at a rate of 2 nm min⁻¹. Substrate heating is found to greatly reduce final surface roughness, increase optical transparency and electrical conductivity. Ge₂Sb₂Te₅ is sputtered from a solid target (Super Conductor Materials, USA) at 25 W d.c., 1.1 mtorr working pressure, 4×10^{-6} torr starting pressure and a rate of 8.5 nm min⁻¹, immediately following the ITO deposition in vacuum.

Optical characterization. Films were measured using a Lambda 1050 spectrometer (Perkin Elmer, USA) in reflection, absorption and transmission using wavelengths between 350 nm and 750 nm. Reflectivity measurements were calibrated against a commercial aluminium standard.

Transfer matrix method. We employ a transfer matrix method described by Heavens²⁹ and successfully applied to various systems^{17,30}. More information can be found in section S5 of the Supplementary Information.

Nanoscale patterning of continuous films and pixels. Patterns are generated by converting a selected greyscale image into a spatially resolved voltage pattern. An Asylum Research MFP-3D atomic force microscope equipped with an ORCA accessory (modified to sustain our higher currents) and a conductive tip (DDESP, Bruker) is used as a nanoscale tip to electrically render several images on both films and pixels as shown in Fig. 2a.

Received 15 January; accepted 13 May 2014.

- Wuttig, M. Phase-change materials—towards a universal memory? *Nature Mater.* **4**, 265–266 (2005).
- Lankhorst, M. H. R., Ketelaars, B. W. S. M. M. & Wolters, R. A. M. Low-cost and nanoscale non-volatile memory concept for future silicon chips. *Nature Mater.* **4**, 347–352 (2005).
- Chen, Y. C. et al. Ultra-thin phase-change bridge memory device using GeSb. *2006 International Electron Devices Meeting (IEDM 2006)* <http://dx.doi.org/10.1109/IEDM.2006.346910> (2006).
- Raoux, S. et al. Phase-change random access memory: a scalable technology. *IBM J. Res. Develop.* **52**, 465–479 (2008).
- Caldwell, M. A., Jeyasingh, R. G. D., Wong, H. S. P. & Milliron, D. J. Nanoscale phase change memory materials. *Nanoscale* **4**, 4382–4392 (2012).
- Ohno, E., Yamada, N., Kurumizawa, T., Kimura, K. & Takao, M. Tegesnau alloys for phase-change type optical disk memories. *Jpn. J. Appl. Phys.* **28**, 1235–1240 (1989).

- Afonso, C. N., Solis, J., Catalina, F. & Kalpouzos, C. Ultrafast reversible phase-change in GeSb films for erasable optical storage. *Appl. Phys. Lett.* **60**, 3123–3125 (1992).
- Wuttig, M. & Yamada, N. Phase-change materials for rewriteable data storage. *Nature Mater.* **6**, 824–832 (2007).
- Kats, M. A., Blanchard, R., Genevet, P. & Capasso, F. Nanometre optical coatings based on strong interference effects in highly absorbing media. *Nature Mater.* **12**, 20–24 (2013).
- Oosthoek, J. L. M. et al. Evolution of cell resistance, threshold voltage and crystallization temperature during cycling of line-cell phase-change random access memory. *J. Appl. Phys.* **110**, 024505 (2011).
- Loke, D. et al. Breaking the speed limits of phase-change memory. *Science* **336**, 1566–1569 (2012).
- Wang, W. J. et al. Engineering grains of Ge₂Sb₂Te₅ for realizing fast-speed, low-power, and low-drift phase-change memories with further multilevel capabilities. *2012 International Electron Devices Meeting (IEDM 2012)* <http://dx.doi.org/10.1109/IEDM.2012.6479143> (2012).
- Xiong, F., Liao, A. D., Estrada, D. & Pop, E. Low-power switching of phase-change materials with carbon nanotube electrodes. *Science* **332**, 568–570 (2011).
- Lee, S. H., Jung, Y. & Agarwal, R. Highly scalable non-volatile and ultra-lowpower phase-change nanowire memory. *Nature Nanotechnol.* **2**, 626–630 (2007).
- Yamada, N. & Matsunaga, T. Structure of laser-crystallized Ge₂Sb_{2-x}Te₅ sputtered thin films for use in optical memory. *J. Appl. Phys.* **88**, 7020–7028 (2000).
- Collings, N., Davey, T., Christmas, J., Chu, D. P. & Crossland, B. The applications and technology of phase-only liquid crystal on silicon devices. *J. Displ. Technol.* **7**, 112–119 (2011).
- Burkhard, G. F., Hoke, E. T. & McGehee, M. D. Accounting for interference, scattering, and electrode absorption to make accurate internal quantum efficiency measurements in organic and other thin solar cells. *Adv. Mater.* **22**, 3293–3297 (2010).
- Bichet, O., Wright, C. D., Samson, Y. & Gidon, S. Local characterization and transformation of phase-change media by scanning thermal probes. *J. Appl. Phys.* **95**, 2360–2364 (2004).
- Satoh, H., Sugawara, K. & Tanaka, K. Nanoscale phase changes in crystalline Ge₂Sb₂Te₅ films using scanning probe microscopes. *J. Appl. Phys.* **99**, 024306 (2006).
- Hamann, H. F., O'Boyle, M., Martin, Y. C., Rooks, M. & Wickramasinghe, K. Ultra-high-density phase-change storage and memory. *Nature Mater.* **5**, 383–387 (2006).
- Bhaskaran, H., Sebastian, A., Pauza, A., Pozidis, H. & Despont, M. Nanoscale phase transformation in Ge₂Sb₂Te₅ using encapsulated scanning probes and retraction force microscopy. *Rev. Sci. Instrum.* **80**, 083701 (2009).
- Lingley, A. R. et al. A single-pixel wireless contact lens display. *J. Microelectromech. Syst.* **21**, 125014 (2011).
- Llordés, A., Garcia, G., Gazquez, J. & Milliron, D. J. Tunable near-infrared and visible-light transmittance in nanocrystal-in-glass composites. *Nature* **500**, 323–326 (2013).
- Zrenner, E. Fighting blindness with microelectronics. *Sci. Transl. Med.* **5**, 210ps216 (2013).
- Nam, S. W. et al. Electrical wind force-driven and dislocation-templated amorphization in phase-change nanowires. *Science* **336**, 1561–1566 (2012).
- Redaelli, A., Pirovano, A., Benvenuti, A. & Lacaita, A. L. Threshold switching and phase transition numerical models for phase change memory simulations. *J. Appl. Phys.* **103**, 111101 (2008).
- Bez, R., Cappelletti, P., Servalli, G. & Pirovano, A. Phase change memories have taken the field. *5th IEEE International Memory Workshop* 13–16, <http://dx.doi.org/10.1109/IMW.2013.6582084> (2013).
- Heikenfeld, J., Drzaic, P., Yeo, J. S. & Koch, T. A critical review of the present and future prospects for electronic paper. *J. Soc. Inf. Displ.* **19**, 129–156 (2011).
- Heavens, O. *Optical Properties of Thin Solid Films* Ch. 4, 46–95 (Dover, 1991).
- Pettersson, L. A. A., Roman, L. S. & Inganas, O. Modeling photocurrent action spectra of photovoltaic devices based on organic thin films. *J. Appl. Phys.* **86**, 487–496 (1999).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. Taylor for scientific discussions related to optical spectroscopy measurements. We are grateful to M. Riede for discussions on the modelling aspects of our study. This research was supported by EPSRC via grant numbers EP/J018783/1, EP/J018694/1 and EP/J00541X/2 as well as the OUP John Fell Fund.

Author Contributions All authors contributed substantially to this work. P.H. and H.B. conceived and designed the experiments. P.H. performed the experiments with input from H. B. and C.D.W. All authors analysed the data. The manuscript was written by P.H. and H.B. with input from C.D.W.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.B. (harish.bhaskaran@materials.ox.ac.uk).

Quantification of dissolved iron sources to the North Atlantic Ocean

Tim M. Conway^{1†} & Seth G. John¹

Dissolved iron is an essential micronutrient for marine phytoplankton, and its availability controls patterns of primary productivity and carbon cycling throughout the oceans^{1,2}. The relative importance of different sources of iron to the oceans is not well known, however, and flux estimates from atmospheric dust, hydrothermal vents and oceanic sediments vary by orders of magnitude. Here we present a high-resolution transect of dissolved stable iron isotope ratios ($\delta^{56}\text{Fe}$) and iron concentrations ([Fe]) along a section of the North Atlantic Ocean. The different iron sources can be identified by their unique $\delta^{56}\text{Fe}$ signatures, which persist throughout the water column. This allows us to calculate the relative contribution from dust, hydrothermal venting and reductive and non-reductive sedimentary release to the dissolved phase. We find that Saharan dust aerosol is the dominant source of dissolved iron along the section, contributing 71–87 per cent of dissolved iron. Additional sources of iron are non-reductive release from oxygenated sediments on the North American margin (10–19 per cent), reductive sedimentary dissolution on the African margin (1–4 per cent) and hydrothermal venting at the Mid-Atlantic Ridge (2–6 per cent). Our data also indicate that hydrothermal vents in the North Atlantic are a source of isotopically light iron, which travels thousands of kilometres from vent sites, potentially influencing surface productivity. Changes in the relative importance of the different iron sources through time may affect interactions between the carbon cycle and climate.

Low Fe concentrations limit nitrogen fixation and photosynthesis over much of the oceans¹, and so both temporal and spatial changes in the supply of dissolved Fe to the oceans markedly affect both marine biogeochemistry and the global carbon cycle². Changes in Fe supply to the oceans from dust aerosols may have a role in interglacial–glacial climate cycling³, and changing future dust sources are likely to have a role in future climate change⁴. Hydrothermal venting may vary less on short timescales, perhaps serving to buffer dissolved Fe concentrations when other sources decrease⁵. Sedimentary margins, especially where pore waters are reducing, are a poorly constrained but potentially important Fe source⁶, the magnitude of which is likely to change with climate-induced variability in global patterns of dissolved oxygen, productivity and organic-rich sediments.

Despite the clear significance of Fe to biogeochemical processes and the global carbon cycle, large uncertainties remain over the relative importance of different Fe sources. Aerosol dust has traditionally been considered a major source of Fe to the surface ocean, especially close to major dust sources⁴. However, constraints on atmospheric Fe fluxes are limited by an understanding of the dissolution of aerosol Fe, with estimates of Fe solubility ranging from 0.1% to 80% (ref. 7). Similar uncertainties complicate estimates of Fe flux from reducing continental-margin sediments, where much of the reduced Fe(II) precipitates on mixing with oxic sea water. Fluxes based on water-column data can be three orders of magnitude lower than fluxes based on porewater data at the same location^{8,9}, and global fluxes predicted from low-oxygen sediments are an order of magnitude higher than observed fluxes from higher-oxygen locations⁹. Equally, although most of the Fe in Fe-rich hydrothermal fluids

precipitates on mixing with sea water, complexation with organic ligands may facilitate the transport of Fe over great distances from vents^{10,11}, leading to uncertainty (2–75%) in the contribution of hydrothermal Fe to the dissolved Fe inventory^{5,11,12}.

Stable isotope ratios of seawater-dissolved Fe ($\delta^{56}\text{Fe}$) provide a novel method^{13–15} of constraining Fe fluxes to the ocean, by ‘fingerprinting’ Fe from different sources with distinct $\delta^{56}\text{Fe}$ (refs 8, 16). $\delta^{56}\text{Fe}$ spans $\sim 4\text{‰}$ in marine environments, with sediments and bulk aerosols at $\sim 0\text{‰}$ (refs 17–19), whereas dissolved Fe influenced by redox chemistry in hydrothermal vent fluids and reducing sediment pore waters is isotopically light^{20–22}. Here, using a new method¹⁵, which allows the determination of $\delta^{56}\text{Fe}$ in sea water with 2σ precision of 0.01–0.3‰ on 1-litre samples as low as 0.05 nmol per kg Fe, we present the first oceanographic section of dissolved $\delta^{56}\text{Fe}$. Samples were collected from 17 stations during the US GEOTRACES North Atlantic GA03 Zonal Transect between Mauritania and Woods Hole (25–37 point profiles, 510 samples; Fig. 1). Patterns in dissolved [Fe] across the section (Fig. 2a) reflect regions of significant Fe input, with higher [Fe] observed in surface waters where dust is deposited; within the Mauritanian oxygen minimum zone (OMZ; 100–1,000 m, 18–35° W; Extended Data Fig. 1); at depths between 800 and 2,500 m extending from the North American shelf as far east as 55° W, corresponding to Upper Labrador Sea Water (ULSW; W. Jenkins, W. Smethie, E. Boyle and G. Cutter, personal communication); and near the TAG hydrothermal vent field within the Mid-Atlantic Ridge (MAR) rift valley (Fig. 3). We found that $\delta^{56}\text{Fe}$ in these four regions was distinct, reflecting unique endmember $\delta^{56}\text{Fe}$ signatures of Fe sourced from dust, from reducing and oxygenated sediments, and from hydrothermal vents. We therefore used Fe isotope mass balance to calculate the relative proportion of Fe from different sources at any point along the high-resolution section.

A striking observation about $\delta^{56}\text{Fe}$ (Fig. 2b) across the section is the dominance of heavier-than-crustal ratios, as high as $+0.6$ to $+0.8\text{‰}$, when the known marine sources are either light^{8,20,22} or close to 0‰ (refs 17, 18, 21). The surface mixed layer right across the section was dominated by $\delta^{56}\text{Fe}$ values of $+0.3$ to $+0.7\text{‰}$, which we attribute to fractionation of bulk

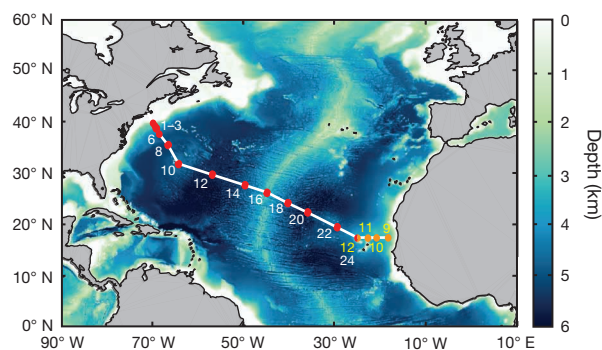


Figure 1 | The cruise track of the US GEOTRACES North Atlantic GA03 Zonal Transect. The 2010 (USGT10, red) and 2011 (USGT11, orange) legs are shown. Samples from 17 numbered stations were analysed in this study.

¹Department of Earth and Ocean Sciences, University of South Carolina, Columbia, South Carolina 29208, USA. [†]Present address: Institute of Geochemistry and Petrology, Eidgenössische Technische Hochschule Zürich, NW D81.4, Clausiusstrasse 25, 8092 Zürich, Switzerland.

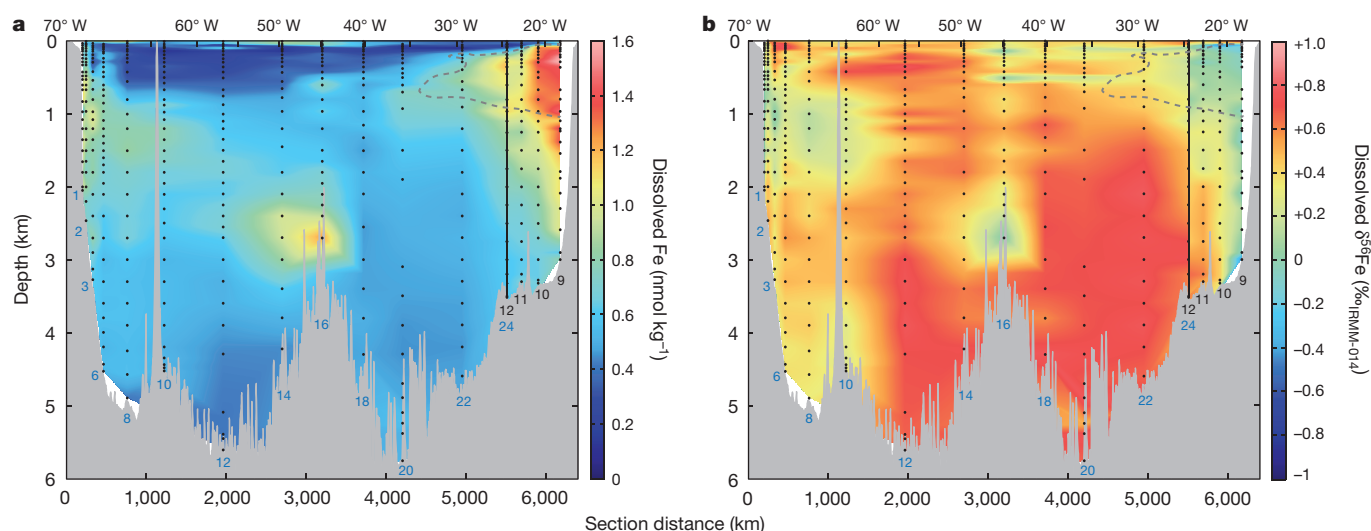


Figure 2 | Dissolved Fe concentration and $\delta^{56}\text{Fe}$ results from the USGT North Atlantic GA03 Zonal Transect. Station numbers are shown in blue for USGT11 and black for USGT10, with sampling depths for each station shown as black dots. The black vertical lines denote the crossover between the two cruises at stations USGT10-12 and USGT11-24. The data set contains 510

samples. **a**, Dissolved Fe concentration (nmol kg^{-1}). **b**, Stable Fe isotope ratio as $\delta^{56}\text{Fe}$ (‰) relative to IRMM-014. The region of low dissolved oxygen concentrations close to West Africa is delineated by the $120 \mu\text{mol kg}^{-1}$ dissolved oxygen contour (grey dashed line). Dissolved oxygen concentrations for the section are shown in Extended Data Figs 1 and 2.

aerosol ($\delta^{56}\text{Fe} \approx 0\text{‰}$; refs 18, 19) during dissolution¹⁶, perhaps as a result of a ligand-mediated process concentrating heavier Fe isotopes with strong Fe-binding ligands²³. To best represent this aerosol Fe source in source fraction calculations, we chose a value of $+0.68\text{‰}$ for the dust endmember on the basis of the heaviest $\delta^{56}\text{Fe}$ values observed in the surface mixed layer close to African and North American dust sources⁷ (Extended Data Figs 2 and 3 and Methods). Attribution of heavy Fe to aerosol dust is also supported by both the signature of the deep Atlantic away from other sources ($+0.68 \pm 0.06\text{‰}$ (mean \pm s.d.)), where dust is expected to dominate, and the light-isotope signature of the water column in regions of the Southern Ocean²⁴ ($+0.1\text{‰}$) and the North Pacific¹⁵ (-0.8 to $+0.1\text{‰}$; Methods) where dust deposition is low. We also note that the overall pattern is inconsistent with the previously proposed surface

biological uptake and deeper regeneration of light Fe isotopes²⁵. Instead, we observed a subsurface excursion to lighter $\delta^{56}\text{Fe}$ values, coincident with a subsurface [Fe] minimum and fluorescence maximum (Extended Data Figs 2 and 3), which hints at the removal of heavy Fe from the dissolved Fe pool associated with biological activity and may be responsible for the slightly lighter surface ocean in the west of the basin, where values of $+0.2$ to $+0.5\text{‰}$ were observed (stations USGT11-10 to 18) within the deep surface mixed layer.

In the eastern basin, within the OMZ under the Mauritanian upwelling zone, light $\delta^{56}\text{Fe}$ signatures, as low as -0.5‰ , are characteristic of Fe(II) released from reductive sediments into the water column^{8,22} (Fig. 2b). We therefore assign a light-endmember value of -2.4‰ to Fe released from reductive African margin sediments, on the basis of the Fe(II) flux

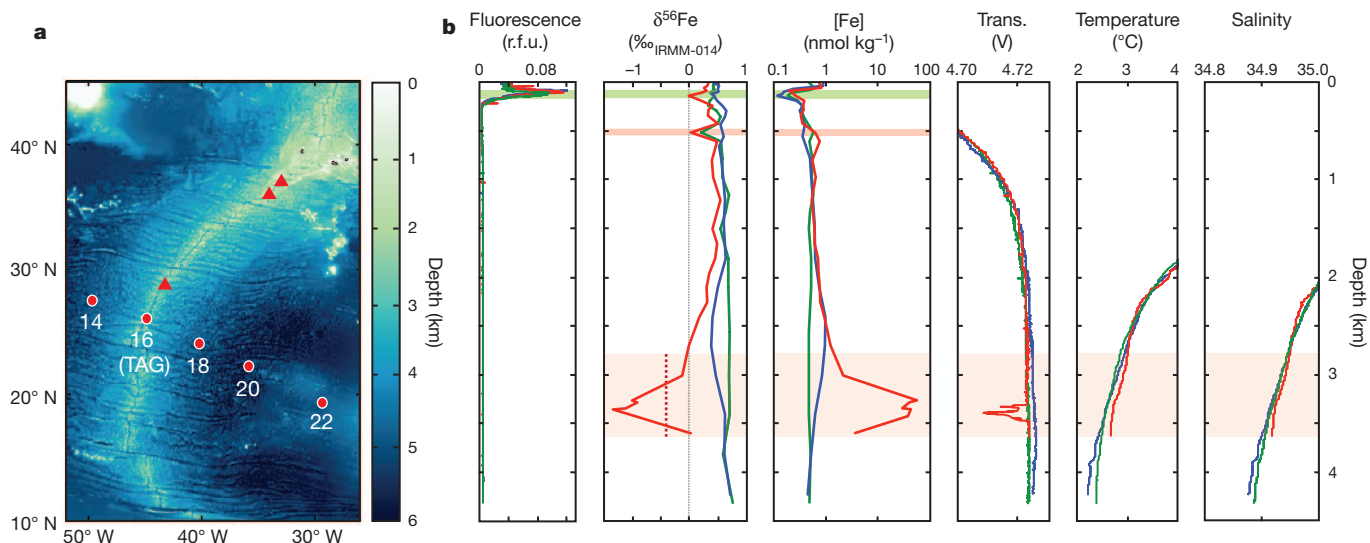


Figure 3 | The influence of hydrothermal venting on dissolved Fe and $\delta^{56}\text{Fe}$. **a**, Location map showing USGT11 stations close to the Mid-Atlantic Ridge, including the TAG hydrothermal site (USGT11-16), and bathymetry. Red triangles show nearby black smoker vents (from the InterRidge website, <http://www.interridge.org/>). **b**, Multiple oceanographic parameters from stations USGT11-14 (blue lines), USGT11-16 (red lines) and USGT11-18 (green lines), together with Fe concentration and $\delta^{56}\text{Fe}$ (relative to IRMM-014). The surface green horizontal band denotes the fluorescence maximum, the red

band (at 0.5 km depth) highlights subsurface hydrothermal Fe, and the broad red band (at 2.6–3.3 km depth) denotes the hydrothermal plume sampled directly at the TAG site, defined by temperature and salinity anomalies. The dotted red line indicates the $\delta^{56}\text{Fe}$ of TAG fluids²⁰, relative to IRMM-014. Transmissometry (Trans., in volts) is shown, with low transmissometry values within the plume indicative of a high concentration of suspended particles. r.f.u., relative fluorescence units.

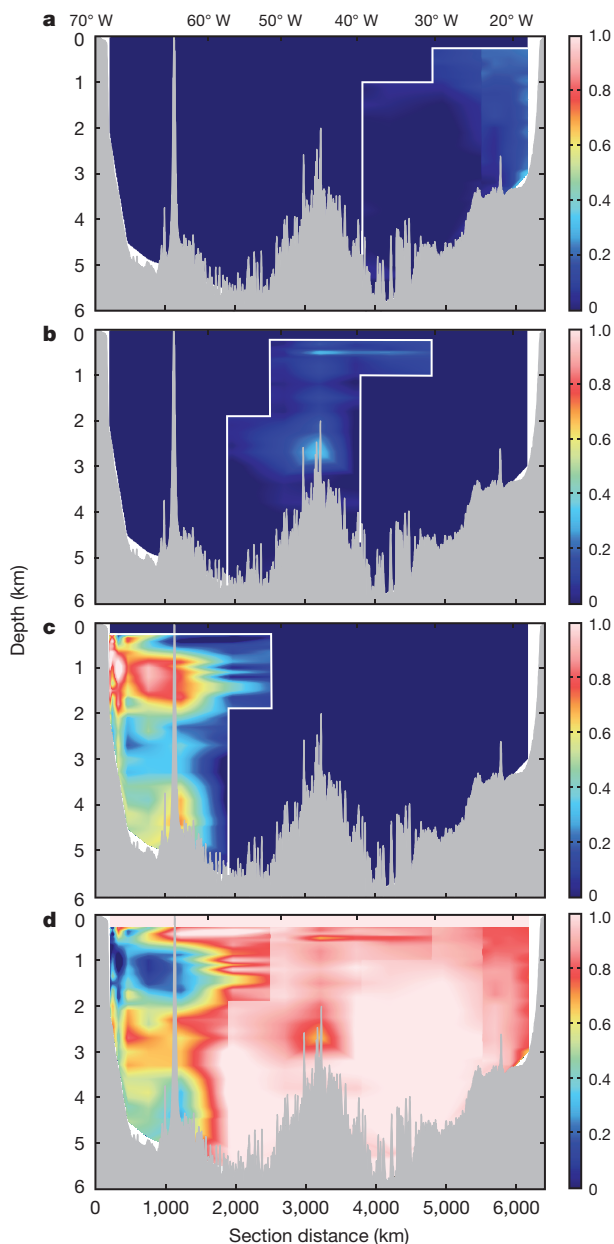


Figure 4 | Quantification of the importance of sources to the dissolved Fe reservoir across the North Atlantic section. The fraction (f) of dissolved Fe across the USGT North Atlantic GA03 Zonal Transect originating from reductive release from sediments (a), the Mid-Atlantic Ridge (b), non-reductive release from sediments (c) or aerosol dust (d), as calculated by two-component isotope mixing (see Methods). See Table 1 and Methods for a sensitivity analysis on the choice of endmember values.

from sediments in the San Pedro basin⁸, and consistent with porewater studies²¹. The observed -0.5‰ $\delta^{56}\text{Fe}$ at 50 m above the sediments therefore reflects a mixture between heavy background aerosol Fe ($+0.68\text{‰}$) and light sedimentary Fe (-2.4‰). Using these endmember values, we quantify the fraction of the high [Fe] near Africa originating from each source. Within the OMZ (200–1,000 m) as far as 32°W , where the majority of the Fe was previously attributed to aerosol dust on the basis of apparent oxygen utilization²⁶, $\delta^{56}\text{Fe}$ -informed calculations indicate that shelf advection contributes 10–30% of the dissolved Fe (Fig. 4). Within a few hundred kilometres of the margin (USGT10–9), 20–30% of the dissolved Fe throughout the water column originates in reductive sediments. Fe from the margin is also transported out $\sim 1,000$ km throughout the deep ocean, outside the OMZ, where dissolved oxygen was $>160\text{ }\mu\text{mol l}^{-1}$.

At the TAG site at the MAR, $\delta^{56}\text{Fe}$ within a hydrothermal plume (2,900–3,600 m; Fig. 3) was as light as -1.35‰ , which we assign as the hydrothermal endmember. The minimum in $\delta^{56}\text{Fe}$ was notably lighter than vent fluids at TAG (-0.4‰ ; ref. 20), consistent with non-quantitative oxidation of Fe^{2+} and precipitation of oxyhydroxides on mixing of Fe-rich vent fluids with ambient sea water²⁰. Hydrothermal Fe was present up to 2 km above the TAG site (USGT11–16), as well as $\sim 1,000$ km to the west at station USGT11–14, but there is no evidence that hydrothermal Fe is transported eastwards to USGT11–18. Both light $\delta^{56}\text{Fe}$ above TAG and a light excursion in $\delta^{56}\text{Fe}$ at 500 m across the centre of the basin (Figs 2 and 3) suggest that there could be a significant hydrothermal contribution to the surface gyre, potentially from shallower vents such as Rainbow (2,220 m)²⁰ or Lucky Strike (1,750 m)¹², close to the Azores, which have similar vent chemistry to that at TAG²⁰. This supply of hydrothermal Fe to the surface ocean could act to buffer the dissolved Fe pool against changing seasonal fluxes of aerosol Fe.

Close to the North American margin, dissolved Fe was distinctly isotopically lighter within the high-[Fe] region corresponding to Upper Labrador Sea Water. Here, a $\delta^{56}\text{Fe}$ of 0 to $+0.2\text{‰}$ indicates Fe originating from a source with a $\delta^{56}\text{Fe}$ signature close to crustal values, either Fe obtained during contact of the water mass with oxygenated margin sediments or from well-oxygenated sediments in the Labrador Sea. The presence of low- $\delta^{56}\text{Fe}$ closer to the surface at margin stations USGT11–1 and USGT11–2, outside ULSW, as well as associated with deep nepheloid layers as far east as Bermuda, indicates that sediments are the likely Fe source. Paired measurements of dissolved and soluble [Fe] (0.2 and $<0.02\text{ }\mu\text{m}$, respectively) near the margin suggest that colloidal iron (0.02–0.2 μm), probably resuspended from sediments, dominates the dissolved Fe ($<0.2\text{ }\mu\text{m}$) pool (J. N. Fitzsimmons, G. G. Carrasco, J. Wu, S. Roshan, M. Hatta, C. I. Measures, T.M.C., S.G.J. and E. A. Boyle, unpublished observations). Regardless of whether dissolved Fe in this region is sourced from 'non-reductive' dissolution of sediments²¹ or from resuspension of sedimentary colloidal Fe, both sources share near-crustal $\delta^{56}\text{Fe}$ values of 0–0.2‰ (refs 17, 21, 22). Using a crustal endmember ($+0.09\text{‰}$; ref. 17) to represent these non-reductive sedimentary sources, we find that non-reductive release from oxygenated sediments is responsible for up to 100% of the [Fe] throughout the region dominated by ULSW (Fig. 4), from the margin to past Bermuda.

Table 1 | Sensitivity analysis of the choice of endmember $\delta^{56}\text{Fe}$ value on calculated percentage source contributions

Aerosol dust endmember $\delta^{56}\text{Fe}$ value (‰)	Aerosol dust (%)			Reductive sediments (%)			Non-reductive sediments (%)			MAR hydrothermal (%)	
	Min.*	Fig. 4†	Max.‡	Min.‡	Fig. 4†	Max.*	Min.‡	Fig. 4†	Max.*	Min./Fig. 4†‡	Max.*
0.40 (no surface)	90.2	94.2	95.4	0.7	0.9	1.1	3.5	4.6	7.9	0.4	0.8
0.61 (no surface)	79.3	85.2	87.3	1.4	2.0	2.4	9.6	11.2	14.9	1.7	3.3
0.65 (no surface)	76.9	83.2	85.5	1.6	2.2	2.7	10.8	12.5	16.2	2.1	4.1
0.68 (surface)	76.2	82.6	84.8	1.7	2.3	2.8	11.2	12.9	16.5	2.3	4.5
0.68 (no surface)	75.1	81.8	84.2	1.8	2.4	3.0	11.6	13.4	17.2	2.4	4.7
0.70 (no surface)	73.9	80.8	83.2	1.9	2.6	3.2	12.2	14.0	17.8	2.7	5.1
0.75 (no surface)	70.9	78.3	80.9	2.3	3.1	3.8	13.5	15.3	19.0	3.3	6.3

The primary sensitivity analysis of the choice of dust member ($+0.61$ to $+0.75\text{‰}$) provides the range of source contributions described in the text. The secondary sensitivity analysis (italics) additionally considers the effect on source percentages from a lighter dust endmember ($+0.4\text{‰}$; see Methods). Sensitivity analyses were performed with -1.82‰ *, -2.40‰ † or -3.45‰ ‡ for the reductive sedimentary endmember, $+0.22\text{‰}$ *, $+0.09\text{‰}$ † or 0.00‰ ‡ for the non-reductive sedimentary endmember, -0.36‰ * or -1.35‰ †‡ for the hydrothermal endmember, and with (surface) or without (no surface) the surface 250 m included. Percentages calculated with the endmember $\delta^{56}\text{Fe}$ values used for Fig. 4 are shown in bold.

By combining the fraction of Fe from various sources (f) obtained from isotope mass-balance with [Fe], we calculated the section-wide contributions of each Fe source (Fig. 4, Table 1 and Methods). We also tested the sensitivity of these calculations to varying the choice of endmember $\delta^{56}\text{Fe}$ values (Table 1). We found aerosol dust to be the dominant source of Fe to the North Atlantic, supplying 71–87% of the Fe along this section, whereas oxygenated sediments supplied 10–19% and dominated the water column in the west. Despite the transport of hydrothermal Fe great distances from vents, which supports the ‘leaky-vent’ hypothesis²⁷, the MAR was not a significant source of Fe across the whole section (~ 1.7 – 6.3%). Here we show definitively that reduced sedimentary Fe is advected throughout the OMZ near Africa, contributing ~ 1.4 – 3.8% of the Fe across the section; this hints at a greater importance of reduced sedimentary Fe worldwide than previously considered, especially in large-scale OMZs such as those close to South America, where distal transport of shelf Fe could be highly sensitive to climate-induced changes in ocean oxygenation. Lower $\delta^{56}\text{Fe}$ in the Southern²⁴ and Pacific¹⁵ oceans suggests that dust aerosol contributes less Fe in these areas, implying regional variability in the sources and processes controlling the global distribution of dissolved Fe.

METHODS SUMMARY

Sea water was collected with the use of clean procedures during the US GEOTRACES North Atlantic GA03 Zonal Transect (USGT10, October to November 2010; USGT11, November to December 2011), into Go-Flow bottles on the GEOTRACES rosette or by surface towfish, and then filtered with 0.2- μm Pall Acropak-200 Supor cartridges into acid-cleaned 1-litre polyethylene bottles. Samples were later acidified to pH ~ 2 with 12 M VWR Aristar ultra HCl and left for several months before being processed. Samples were analysed at the University of South Carolina by Neptune multiple-collector inductively coupled plasma mass spectrometry (MC-ICP-MS) after the addition of a ^{57}Fe – ^{58}Fe double spike, in accordance with published methods¹⁵. [Fe] in each sample was calculated by isotope dilution, and 2% error was assigned to account for weighing and pipetting error. Determination of [Fe] in reference standards from the SArFe site in the North Pacific (30°N , 140°W), analysed alongside samples, showed excellent agreement with 2013 consensus values¹⁵. Stable isotope ratios are all reported relative to isotope standard IRRM-014:

$$\delta^{56}\text{Fe}(\text{‰}) = 1,000 \times \left(\left(\frac{{}^{56}\text{Fe}}{{}^{54}\text{Fe}} \right)_{\text{sample}} / \left(\frac{{}^{56}\text{Fe}}{{}^{54}\text{Fe}} \right)_{\text{IRMM-014}} - 1 \right) \quad (1)$$

We express error (see Supplementary Data) as the combined 2σ standard internal error of samples and standards, as described previously¹⁵. The distribution of [Fe] and $\delta^{56}\text{Fe}$ was linearly interpolated with MatLab, first vertically then horizontally. For fraction calculations (Fig. 4 and Table 1), we subdivided the section (below 250 m) into three boxes (Fig. 4), on the basis of the large-scale features of $\delta^{56}\text{Fe}$ and [Fe]. Within each box the dissolved Fe was then described by mixing between two Fe sources with distinct $\delta^{56}\text{Fe}$ signatures. Fe from western non-reductive sediments ($f_{\text{non-red}} + 0.09\%$), from the MAR ($f_{\text{hyd}} - 1.35\%$) or from eastern reductive sediments ($f_{\text{red}} - 2.40\%$) was mixed with a dust endmember ($f_{\text{dust}} + 0.68\%$) using two-component mixing:

$$\delta^{56}\text{Fe}_{\text{dissolved}} = (f_1 \times \delta^{56}\text{Fe}_1) + (f_2 \times \delta^{56}\text{Fe}_2) \quad (2)$$

where f and $\delta^{56}\text{Fe}$ denote the fraction and isotopic signature from each source. A sensitivity analysis of endmember choice is discussed in Methods and is shown in Table 1.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 January; accepted 9 May 2014.

Published online 2 July 2014.

1. Moore, J. K., Doney, S. C., Glover, D. M. & Fung, I. Y. Iron cycling and nutrient-limitation patterns in surface waters of the World Ocean. *Deep Sea Res. II* **49**, 463–507 (2002).
2. Boyd, P. W. & Ellwood, M. J. The biogeochemical cycle of iron in the ocean. *Nature Geosci.* **3**, 675–682 (2010).

3. Ridgwell, A. & Kohfeld, K. E. Glacial-interglacial variability in atmospheric CO_2 . *Geophys. Monogr. Ser.* **17**, 251–286 (2009).
4. Jickells, T. D. *et al.* Global iron connections between desert dust, ocean biogeochemistry, and climate. *Science* **308**, 67–71 (2005).
5. Tagliabue, A. *et al.* Hydrothermal contribution to the oceanic dissolved iron inventory. *Nature Geosci.* **3**, 252–256 (2010).
6. Jeandel, C. *et al.* Ocean margins: the missing term in oceanic element budgets? *Eos* **92**, 217 (2011).
7. Mahowald, N. M. *et al.* Atmospheric global dust cycle and iron inputs to the ocean. *Glob. Biogeochem. Cycles* **19**, GB4025 (2005).
8. John, S. G., Mendez, J., Moffett, J. & Adkins, J. The flux of iron and iron isotopes from San Pedro Basin sediments. *Geochim. Cosmochim. Acta* **93**, 14–29 (2012).
9. Elrod, V. A., Berelson, W. M., Coale, K. H. & Johnson, K. S. The flux of iron from continental shelf sediments: a missing source for global budgets. *Geophys. Res. Lett.* **31**, L12307 (2004).
10. Toner, B. M. *et al.* Preservation of iron(II) by carbon-rich matrices in a hydrothermal plume. *Nature Geosci.* **2**, 197–201 (2009).
11. Saito, M. A. *et al.* Slow-spreading submarine ridges in the South Atlantic as a significant oceanic iron source. *Nature Geosci.* **6**, 775–779 (2013).
12. Carazzo, G., Jellinek, A. M. & Turchyn, A. V. The remarkable longevity of submarine plumes: implications for the hydrothermal input of iron to the deep-ocean. *Earth Planet. Sci. Lett.* **382**, 66–76 (2013).
13. John, S. G. & Adkins, J. F. Analysis of dissolved iron isotopes in seawater. *Mar. Chem.* **119**, 65–76 (2010).
14. Lacan, F. *et al.* High-precision determination of the isotopic composition of dissolved iron in iron depleted seawater by double spike multicollector-ICPMS. *Anal. Chem.* **82**, 7103–7111 (2010).
15. Conway, T. M., Rosenberg, A. D., Adkins, J. F. & John, S. G. A new method for precise determination of iron, zinc and cadmium stable isotope ratios in seawater by double-spike mass spectrometry. *Anal. Chim. Acta* **793**, 44–52 (2013).
16. John, S. G. & Adkins, J. F. The vertical distribution of iron stable isotopes in the North Atlantic near Bermuda. *Glob. Biogeochem. Cycles* **26**, GB2034 (2012).
17. Beard, B. L. & Johnson, C. M. Iron isotope constraints on Fe cycling and mass balance in oxygenated Earth oceans. *Geology* **31**, 629–632 (2003).
18. Waelles, M., Baker, A. R., Jickells, T. & Hoogewerff, J. Global dust teleconnections: aerosol iron solubility and stable isotope composition. *Environ. Chem.* **4**, 233–237 (2007).
19. Mead, C., Herckes, P., Majestic, B. J. & Anbar, A. D. Source apportionment of aerosol iron in the marine environment using iron isotope analysis. *Geophys. Res. Lett.* **40**, 5722–5727 (2013).
20. Severmann, S. *et al.* The effect of plume processes on the Fe isotope composition of hydrothermally derived Fe in the deep ocean as inferred from the Rainbow vent site, Mid-Atlantic Ridge, $36^\circ 14' \text{N}$. *Earth Planet. Sci. Lett.* **225**, 63–76 (2004).
21. Hornok, W. B., John, S. G., Conway, T. M. & Mills, R. A. Distinct iron isotopic signatures and supply from marine sediment dissolution. *Nature Commun.* **4**, 2143 (2013).
22. Hornok, W. B., Severmann, S., Mills, R. A., Statham, P. J. & Fones, G. R. Pore-fluid Fe isotopes reflect the extent of benthic Fe redox recycling: evidence from continental shelf and deep-sea sediments. *Geology* **37**, 751–754 (2009).
23. Iliina, S. M. *et al.* Extreme iron isotope fractionation between colloids and particles of boreal and temperate organic-rich waters. *Geochim. Cosmochim. Acta* **101**, 96–111 (2013).
24. Lacan, F. *et al.* Measurement of the isotopic composition of dissolved iron in the open ocean. *Geophys. Res. Lett.* **35**, L24610 (2008).
25. Radic, A., Lacan, F. & Murray, J. W. Iron isotopes in the seawater of the equatorial Pacific Ocean: new constraints for the oceanic iron cycle. *Earth Planet. Sci. Lett.* **306**, 1–10 (2011).
26. Fitzsimmons, J. N., Zhang, R. & Boyle, E. A. Dissolved iron in the tropical North Atlantic Ocean. *Mar. Chem.* **154**, 87–99 (2013).
27. Toner, B. M., Marcus, M. A., Edwards, K. J., Rouxel, O. & German, C. R. Measuring the form of iron in hydrothermal plume particles. *Oceanography* **25**, 209–212 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the Captain, crew and GEOTRACES sampling team on RV *Knorr* on both USGT10 and USGT11; A. Rosenberg for technical assistance; J. Wu, R. Middag, K. Bruland and P. Sedwick for supplying concentration data that informed double-spike calculations; and C. Hayes for calibrating transmissometry data. We thank W. Jenkins, W. Smethie, E. Boyle and G. Cutter for water mass observations. The Ocean Data Facility supplied temperature, PSS-78 salinity, fluorescence, transmissometry, dissolved oxygen and nutrient data from both cruises. Figures were created with Matlab and Adobe Illustrator. This study was funded by National Science Foundation grant OCE-1131387.

Author Contributions The paper was written by T.M.C. and S.G.J. Samples were processed and analysed by T.M.C.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.M.C. (conway.tm@gmail.com).

Strong contributions of local background climate to urban heat islands

Lei Zhao^{1,2}, Xuhui Lee^{1,2}, Ronald B. Smith³ & Keith Oleson⁴

The urban heat island (UHI), a common phenomenon in which surface temperatures are higher in urban areas than in surrounding rural areas, represents one of the most significant human-induced changes to Earth's surface climate^{1,2}. Even though they are localized hotspots in the landscape, UHIs have a profound impact on the lives of urban residents, who comprise more than half of the world's population³. A barrier to UHI mitigation is the lack of quantitative attribution of the various contributions to UHI intensity⁴ (expressed as the temperature difference between urban and rural areas, ΔT). A common perception is that reduction in evaporative cooling in urban land is the dominant driver of ΔT (ref. 5). Here we use a climate model to show that, for cities across North America, geographic variations in daytime ΔT are largely explained by variations in the efficiency with which urban and rural areas convect heat to the lower atmosphere. If urban areas are aerodynamically smoother than surrounding rural areas, urban heat dissipation is relatively less efficient and urban warming occurs (and vice versa). This convection effect depends on the local background climate, increasing daytime ΔT by

3.0 ± 0.3 kelvin (mean and standard error) in humid climates but decreasing ΔT by 1.5 ± 0.2 kelvin in dry climates. In the humid eastern United States, there is evidence of higher ΔT in drier years. These relationships imply that UHIs will exacerbate heatwave stress on human health in wet climates where high temperature effects are already compounded by high air humidity^{6,7} and in drier years when positive temperature anomalies may be reinforced by a precipitation–temperature feedback⁸. Our results support albedo management as a viable means of reducing ΔT on large scales^{9,10}.

The conversion of natural land to urban land causes several notable perturbations to the Earth's surface energy balance. Reduction of evaporative cooling is generally thought to be the dominant factor contributing to UHI. Anthropogenic heat release is an added energy input to the energy balance and should increase the surface temperature. Energy input by solar radiation will also increase if albedo is reduced in the process of land conversion. Buildings and other artificial materials can store more radiation energy in the daytime than can natural vegetation and soil; release of the stored energy at night contributes to night-time

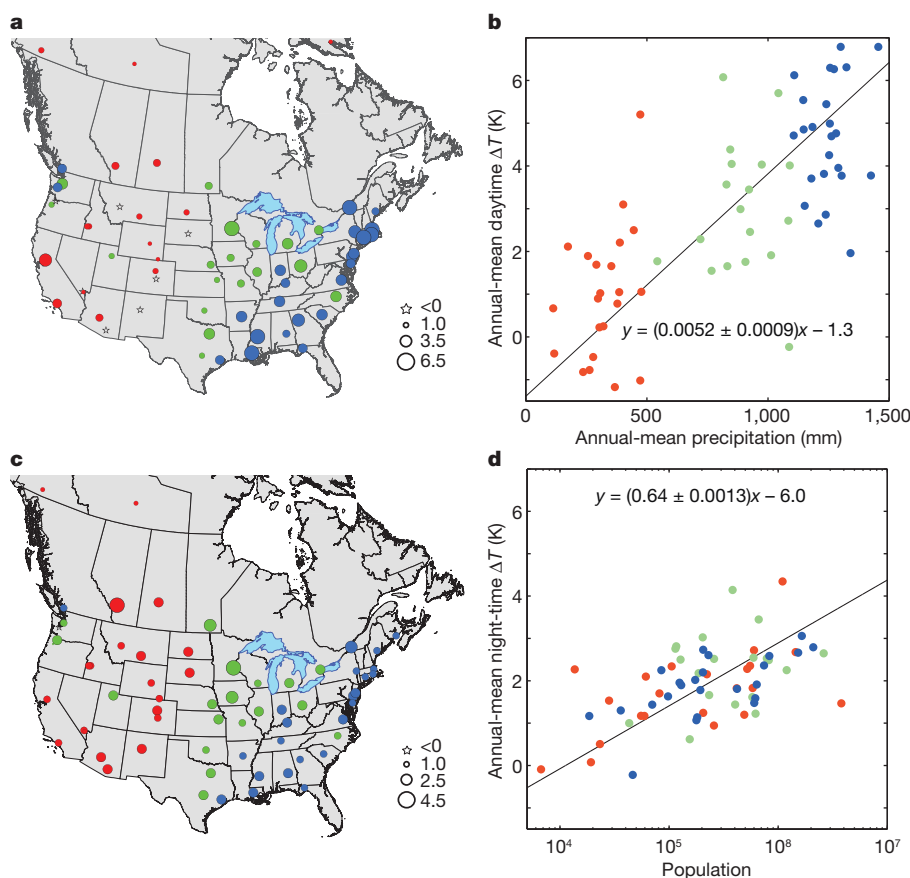


Figure 1 | Precipitation and population influences on MODIS-derived annual-mean UHI intensity. **a**, Map of daytime UHI (shown in K by symbol type/size). **b**, Dependence of daytime UHI on precipitation ($r = 0.74$, $P < 0.001$). **c**, Map of night-time UHI. **d**, Dependence of night-time UHI on population ($r = 0.54$, $P < 0.001$). Red, green and blue symbols denote cities with annual mean precipitations less than 500 mm, between 500 and 1,100 mm, and over 1,100 mm, respectively. Lines in **b** and **d** are linear regression fits to the data. Parameter bounds for the regression slope are the 95% confidence interval.

¹Yale-NUIST Center on Atmospheric Environment, Nanjing University of Information Science and Technology, Nanjing 210044, China. ²School of Forestry and Environmental Studies, Yale University, New Haven, Connecticut 06511, USA. ³Department of Geology and Geophysics, Yale University, New Haven, Connecticut 06511, USA. ⁴National Center for Atmospheric Research, Boulder, Colorado 80305, USA.

UHI. Finally, energy redistribution through convection between the surface and the atmospheric boundary layer can either increase or reduce ΔT , depending on whether the efficiency of convection over urban land is suppressed or enhanced relative to that over adjacent rural land. Although these concepts have been known for some time^{11–13}, a quantitative understanding of their roles across different climate regimes remains elusive⁴.

The climatic context can be understood by posing the following question in a thought experiment: if two cities are built identically in terms of morphological and anthropogenic aspects but in different climates, will they have the same ΔT ? The answer depends on time of day according to observations of surface temperature by the NASA MODIS satellite. For 65 selected cities in North America, the annual-mean midnight ΔT (surface temperature of urban core pixels minus that of rural pixels) is positively correlated with the logarithm of population (correlation coefficient, $r = 0.54$; confidence level, $P < 0.001$; Fig. 1d), but is invariant with climate, showing a statistically insignificant correlation with precipitation ($r = 0.05$, $P = 0.70$; Extended Data Fig. 1), solar radiation ($r = 0.15$, $P > 0.20$) and air temperature ($r = 0.20$, $P > 0.10$). However, the annual-mean midday ΔT is strongly correlated with precipitation ($r = 0.74$, $P < 0.001$; Fig. 1b) and has a weaker statistical dependence on population size than does the night-time ΔT ($r = 0.27$, $P = 0.027$; Extended Data Fig. 1). The night-time ΔT shows little spatial coherence (Fig. 1c), but the daytime ΔT has a discernible spatial pattern that follows precipitation gradients across the continent (Fig. 1a). Twenty-four of the cities are located in the humid southeast United States, which coincides roughly with the Köppen–Geiger temperate climate zone (Fig. 2a). Their daytime annual-mean ΔT is on average 3.9 K and is 3.3 K higher than that of the 15 cities in the dry region (Fig. 2d, e). By comparison, the night-time ΔT differs by 0.1 K between the two groups ($P > 0.60$; Fig. 2f, g). These results are in broad agreement with previous remote-sensing studies on UHI across biophysical and developmental gradients^{14–17}.

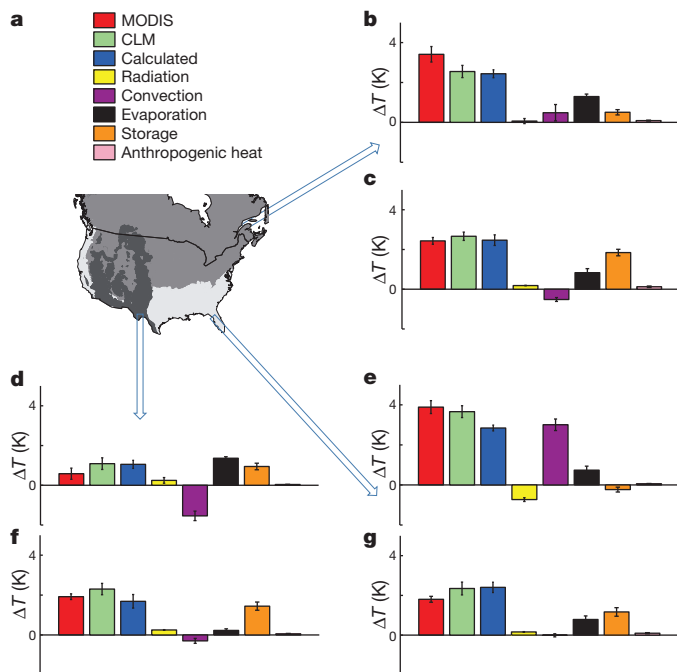


Figure 2 | Attribution of UHI intensity in three Köppen–Geiger climate zones. **a**, Map of climate zones: white, mild temperate/mesothermal climate; grey, continental/microthermal climate; dark grey, dry climate. **b**, **d**, **e**, Daytime values of MODIS and modelled ΔT and its component contributions in each of the three zones (see arrows). **c**, **f**, **g**, Night-time values in each of the three zones (see arrows). Green bars denote model-predicted ΔT and blue bars denote UHI intensity calculated as the sum of the component contributions. Error bars, 1 s.e. for each climate zone.

At first glance, the relationship with precipitation (Fig. 1b) seems consistent with the hypothesis that reduction in evaporative cooling in urban land is the main driver of daytime ΔT , because the denser vegetation in wet climate regions has a higher evaporation rate than the vegetation in dry climates. However, our model-based analysis does not support such an interpretation. In the model domain, ΔT is a perturbation signal to the surface temperature caused by biophysical contrast between rural and urban land units in the same model grid cell¹⁸. This signal is further decomposed, using the method described in ref. 19, into contributions from changes in radiation balance, evaporation, convection efficiency and heat storage, and from anthropogenic heat addition (Fig. 2). The credibility of the model is supported by the reasonable agreement of the modelled ΔT with the MODIS ΔT ($r = 0.31$, $P < 0.02$ for daytime; $r = 0.30$, $P < 0.025$ for night time) and by its accurate depiction of the relationship between night-time ΔT and albedo (Extended Data Fig. 4). Furthermore, the model has reproduced the observed positive correlation between the daytime ΔT and precipitation (Fig. 3a).

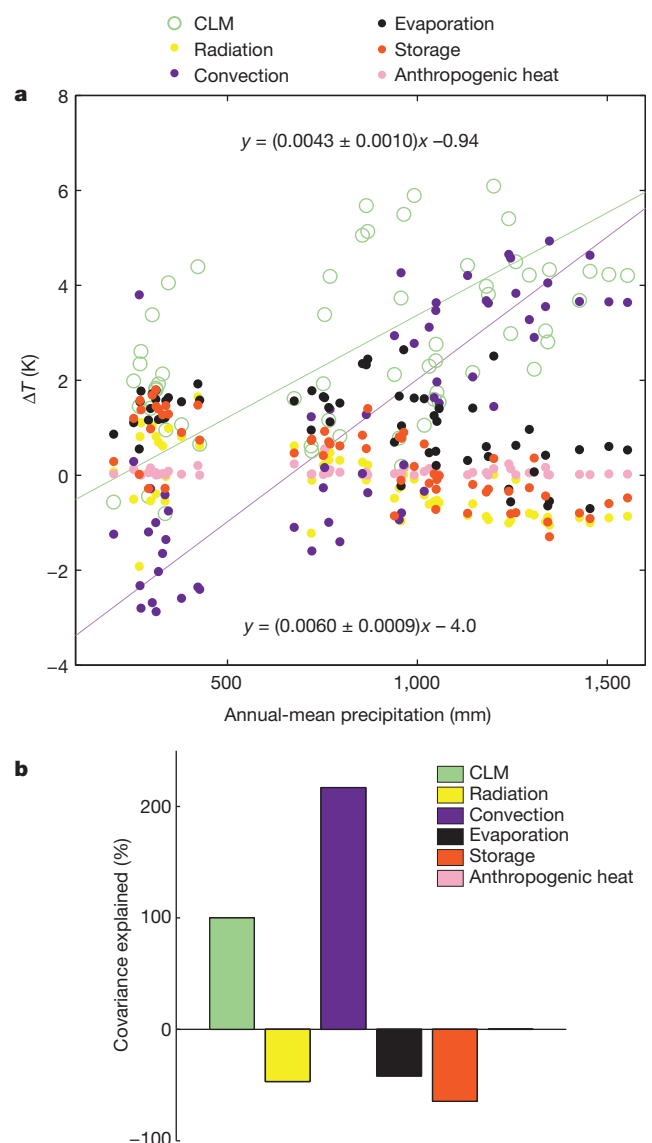


Figure 3 | Relationship between model-predicted daytime ΔT and precipitation among the cities. **a**, Correlation of ΔT and the individual biophysical components with annual-mean precipitation. Lines are linear regression fits to the corresponding data. Parameter bounds for the regression slope are the 95% confidence interval. **b**, ΔT –precipitation covariance explained by different biophysical factors. Note that the covariance explained by the anthropogenic heat term is negligibly small.

We find that it is the changes in convection efficiency (associated with aerodynamic resistance changes), rather than those in evapotranspiration, that control the daytime ΔT –precipitation spatial covariance among the cities (Fig. 3). In the humid climate (the Köppen–Geiger temperate climate zone), convection is less efficient at dissipating heat from urban land than from rural land, and the associated temperature increase is 3.0 ± 0.3 K, which dominates the overall ΔT (Fig. 2e). At these locations, the rural land is in general densely vegetated, owing to ample precipitation, and is aerodynamically rough. Quantitatively, this difference is manifested in a lower aerodynamic resistance to sensible heat diffusion in the rural land (39 s m^{-1}) than in the urban land (62 s m^{-1}). Measured in terms of aerodynamic resistance, urbanization has reduced the convection efficiency by 58%.

The opposite occurs in the dry climate zone, where urban land is rougher than rural land and has enhanced convection efficiency. The result is actually a cooling effect (Fig. 2d). In this zone, the urban landscape has lower aerodynamic resistance (53 s m^{-1}) than the adjacent rural land (66 s m^{-1}), which is typically inhabited by vegetation of low stature such as shrubs, sagebrushes and grasses. On average, the urban land is about 20% more efficient in removing heat from the surface by convection than is the rural land. The average cooling signal is -1.5 ± 0.2 K. In a few of the cities, convection is much more efficient than in the surrounding natural land, such that ΔT becomes negative (Figs 1a and 3a). It has been suggested that negative ΔT , a phenomenon known as ‘urban heat sink’, arises from evaporative cooling of trees and lawns planted in the city^{15–17}. Our explanation seems more logical, because the MODIS urban temperature comes from the urban core pixels with negligible amounts of vegetation cover (enhanced vegetation index, <0.18) and the urban land unit in the climate model is completely free of vegetation. An analogous situation exists in a semi-arid plantation forest where trees serve as efficient ‘heat convectors’, leading to a lower surface temperature than in the adjacent smoother shrub land²⁰.

At night, release of the stored heat is the dominant contributor to ΔT across all three climate zones (Fig. 2c, f and g). The dependence on population size (Fig. 1d), which is an indicator of the city’s horizontal dimension, can be understood in light of these results. At night, the released heat is trapped in a very shallow atmospheric boundary layer. As

air moves across the urban land, it will accumulate more heat with increasing travel distance. Having a longer upwind fetch, that is, a longer distance between the upwind edge of the city and the point of observation, the centre of a larger city should experience stronger warming¹¹.

There is some evidence of precipitation control on interannual variability in the daytime ΔT for individual cities. For each city, we have calculated the linear regression slope of the annual daytime ΔT against the annual precipitation, and we refer to it as the temporal sensitivity to precipitation. Both the MODIS and the model data show a negative dependence of the sensitivity on site mean precipitation (Fig. 4b, d). Twenty-four cities have annual-mean precipitation exceeding 1,100 mm. According to the model, all of these cities, which are mostly distributed in the eastern United States, have negative temporal sensitivity (Fig. 4c), meaning higher ΔT in drier years. The mean temporal sensitivity of this group of cities is $-0.0021 \text{ K mm}^{-1}$. The MODIS results are less consistent because of shorter data records, showing negative sensitivity for 42% of them (Fig. 4a, b).

To gain further insight into the interannual variability, we have compared the daytime ΔT –precipitation correlations for Billings in Montana (annual-mean precipitation, 353 mm) and Richmond in Virginia (1,183 mm). We choose these two cities because they have nearly the same morphological and biophysical specifications (Extended Data Table 1) and therefore are essentially identical in the model world. The sensitivity to precipitation is positive at Billings and negative at Richmond (Extended Data Fig. 2). In contrast to the spatial variations across North America (Fig. 3b), the ΔT interannual variability shown here is driven primarily by changes in surface evaporation (Extended Data Fig. 3).

Our results can be interpreted in the context of heatwave climatology. A measure of heatwave intensity is the degree of deviation, in multiples of standard deviation (North American mean value, $\sigma \approx 0.6$ K) of summertime temperature from the climatological mean²¹. For example, the 2003 European heatwave⁸ is a rare event measured at 5σ . These statistical considerations are based on regional background climatology. Being an additional anomaly on this background condition, UHI will aggravate heat stress on human health. In the southeast United States, where the heat stress is already amplified by high air humidity⁷, the daytime ΔT is equivalent to 7σ (Fig. 2e). The situation may be further

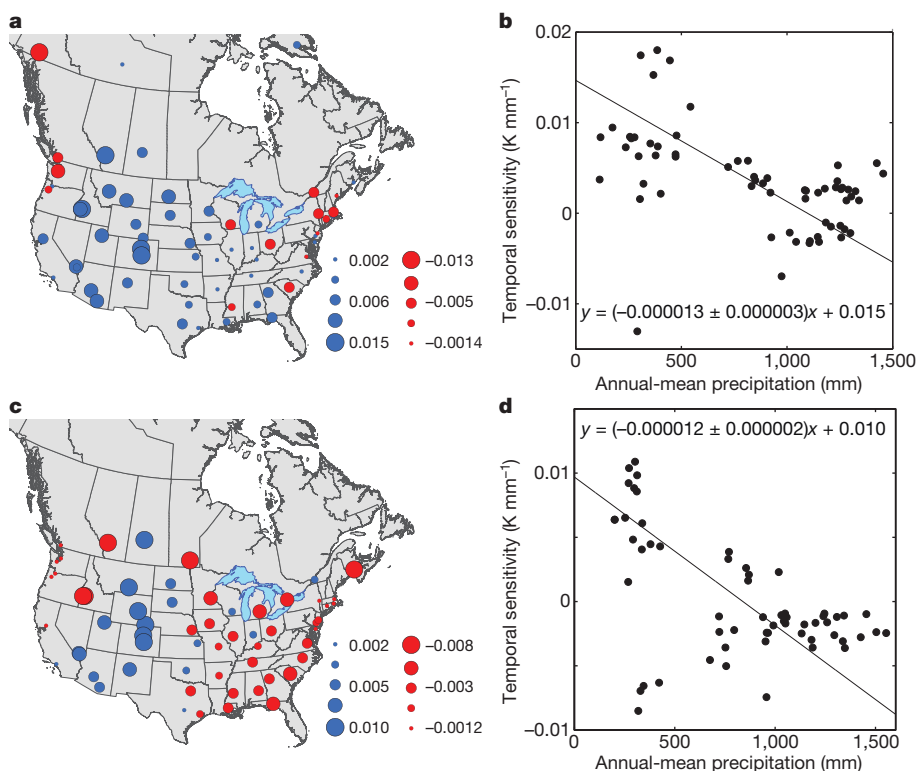


Figure 4 | Temporal sensitivity of UHI intensity to precipitation. a, c, Map of the temporal sensitivities (shown in K mm^{-1} by symbol size) according to MODIS (a) and the climate model (c). b, d, Dependence of MODIS (b) and model-predicted (d) temporal sensitivity on annual mean precipitation. The outlier city in the MODIS panels is Whitehorse in Yukon. The four outlier cities in the model panels are Boise and Nampa in Idaho, Winnipeg in Manitoba and Calgary in Alberta. Lines in b and d are linear regression fits to the data. Parameter bounds for the regression slope are the 95% confidence interval.

worsened in drier years when the positive temperature anomaly is likely to increase owing to a precipitation–temperature feedback⁸. Empirical evidence exists for such synergistic effects²². Using the temporal sensitivity of -0.0021 K mm^{-1} , a 500 mm reduction in the annual precipitation corresponds to an increase in the daytime ΔT by 1.1 K, or $\sim 2\sigma$. We caution that these numbers represent the upper bound of the UHI-added stress because UHI intensity at the screen height²³ (the height of air temperature observation at a standard weather station) and under all-sky conditions should be smaller than our ΔT , which is for clear skies and for the surface. However, summertime ΔT is generally larger than annual ΔT (refs 13–17, 24).

The health impact of heatwaves is one factor that motivates the growing efforts to mitigate UHI. According to our results, a strategy that focuses on reducing anthropogenic heat would bring virtually no relief, but this might be because of the primitive anthropogenic heat scheme in the model¹⁸. Managing the convection efficiency or heat storage of urban land does not seem viable, even though these are large contributors to ΔT , because it would require fundamental changes to the urban morphology, such as a city-wide increase in building height. However, efforts to increase urban albedo have the promise of producing measurable results on a large scale. For the cities in the southern United States, the reduction of net radiation loading amounts to a daytime cooling effect of 0.7 K (Fig. 2e). In the model, this reduction is caused by the fact that these cities have an average albedo that is 0.06 higher than the surrounding rural land. This albedo difference is modest, considering that phasing in reflective roofs in Chicago²⁵ has already increased the city-wide albedo by ~ 0.02 and that some cool-roof implementations¹⁰ aim to increase the urban–rural albedo contrast by as much as 0.6. Albedo increases have little direct effect on the night-time UHI (Fig. 2g) but may have an indirect cooling benefit through the reduction in the daytime heat storage and, therefore, less heat release from storage at night^{16,26,27}. The negative correlation between the night-time ΔT and urban–rural albedo contrast¹⁶ (Extended Data Fig. 4) can be viewed as empirical evidence of this indirect benefit.

METHODS SUMMARY

MODIS data. We calculated the annual-mean ΔT using the MODIS-Aqua eight-day composite land surface temperature from 2003 to 2012. The night-time and daytime ΔT were determined at 1:30 and 13:30 local time, respectively. Data were collected at 65 cities distributed across the United States and Canada. For each city, we paired pixels in the centre of the city with those outside the city to determine ΔT . **Climate model.** We used the Community Earth System Model²⁸ to simulate UHI. The model grid cell consists of urban and rural land units whose surface energy balance variables are calculated using a single-layer urban surface parameterization and a standard land surface scheme, respectively^{18,29}. We ran the model for 33 yr of simulation time from 1972 to 2004 after a 60 yr spin-up. The forcing data are an atmospheric reanalysis product validated against various observations³⁰. The simulation was conducted at the finest resolution supported by the model (0.23° longitude \times 0.31° latitude) to resolve individual cities. The surface skin temperature was determined from the emitted long-wave radiation. To be consistent with the MODIS observations, we used the modelled data at 1:00 and 13:00 local time and under clear-sky conditions to compute the annual ΔT and to perform the surface energy balance analysis.

Attribution of UHI. Attribution of UHI is accomplished by a surface energy balance analysis. The total ΔT is partitioned, using the method of ref. 19, into contributions from the differences, between the urban and rural land units, in surface radiation balance, convection efficiency, evapotranspiration and heat storage, and from anthropogenic heat addition. The perturbation to the radiation balance results mainly from albedo contrast and also includes a minor part associated with surface emissivity change. The analysis was done separately for daytime and night time using the relevant forcing and prognostic model variables.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 January; accepted 7 May 2014.

1. Kalnay, E. & Cai, M. Impact of urbanization and land-use change on climate. *Nature* **423**, 528–531 (2003).

2. Zhou, L. M. *et al.* Evidence for a significant urbanization effect on climate in China. *Proc. Natl Acad. Sci. USA* **101**, 9540–9544 (2004).
3. Grimm, N. B. *et al.* Global change and the ecology of cities. *Science* **319**, 756–760 (2008).
4. Voogt, J. A. & Oke, T. R. Thermal remote sensing of urban climates. *Remote Sens. Environ.* **86**, 370–384 (2003).
5. Taha, H. Urban climates and heat islands: albedo, evapotranspiration, and anthropogenic heat. *Energy Build.* **25**, 99–103 (1997).
6. Fischer, E. M. & Schär, C. Consistent geographical patterns of changes in high-impact European heatwaves. *Nature Geosci.* **3**, 398–403 (2010).
7. Smith, T. T., Zaitchik, B. F. & Gohlke, J. M. Heat waves in the United States: definitions, patterns and trends. *Clim. Change* **118**, 811–825 (2013).
8. Schär, C. *et al.* The role of increasing temperature variability in European summer heatwaves. *Nature* **427**, 332–336 (2004).
9. Akbari, H., Menon, S. & Rosenfeld, A. Global cooling: increasing world-wide urban albedos to offset CO₂. *Clim. Change* **94**, 275–286 (2009).
10. Georgescu, M., Moustauoui, M., Mahalov, A. & Dudhia, J. Summer-time climate impacts of projected megapolitan expansion in Arizona. *Nature Clim. Change* **3**, 37–41 (2013).
11. Oke, T. R. The energetic basis of the urban heat-island. *Q. J. R. Meteorol. Soc.* **108**, 1–24 (1982).
12. Grimmond, S. Urbanization and global environmental change: local effects of urban warming. *Geogr. J.* **173**, 83–88 (2007).
13. Arnfield, A. J. Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island. *Int. J. Climatol.* **23**, 1–26 (2003).
14. Roth, M., Oke, T. R. & Emery, W. J. Satellite-derived urban heat islands from 3 coastal cities and the utilization of such data in urban climatology. *Int. J. Remote Sens.* **10**, 1699–1720 (1989).
15. Imhoff, M. L., Zhang, P., Wolfe, R. E. & Bounoua, L. Remote sensing of the urban heat island effect across biomes in the continental USA. *Remote Sens. Environ.* **114**, 504–513 (2010).
16. Peng, S. S. *et al.* Surface urban heat island across 419 global big cities. *Environ. Sci. Technol.* **46**, 696–703 (2012).
17. Clinton, N. & Gong, P. MODIS detected surface urban heat islands and sinks: global locations and controls. *Remote Sens. Environ.* **134**, 294–304 (2013).
18. Oleson, K. Contrasts between urban and rural climate in CCSM4 CMIP5 climate change scenarios. *J. Clim.* **25**, 1390–1412 (2012).
19. Lee, X. *et al.* Observed increase in local cooling effect of deforestation at higher latitudes. *Nature* **479**, 384–387 (2011).
20. Rotenberg, E. & Yakir, D. Contribution of semi-arid forests to the climate system. *Science* **327**, 451–454 (2010).
21. Hansen, J., Sato, M. & Ruedy, R. Perception of climate change. *Proc. Natl Acad. Sci. USA* **109**, E2415–E2423 (2012).
22. Li, D. & Bou-Zeid, E. Synergistic interactions between urban heat islands and heat waves: the impact in cities is larger than the sum of its parts. *J. Appl. Meteorol. Climatol.* **52**, 2051–2064 (2013).
23. Gallo, K. P., Adegoke, J. O., Owen, T. W. & Elvidge, C. D. Satellite-based detection of global urban heat-island temperature influence. *J. Geophys. Res.* **107**, 4776 (2002).
24. Tran, H., Uchiyama, D., Ochi, S. & Yasuoka, Y. Assessment with satellite data of the urban heat island effects in Asian mega cities. *Int. J. Appl. Earth Obs. Geoinf.* **8**, 34–48 (2006).
25. Mackey, C. W., Lee, X. & Smith, R. B. Remotely sensing the cooling effects of city scale efforts to reduce urban heat island. *Build. Environ.* **49**, 348–358 (2012).
26. Oleson, K. W., Bonan, G. B., Feddesma, J. & Vertenstein, M. An urban parameterization for a global climate model. Part II: sensitivity to input parameters and the simulated urban heat island in offline simulations. *J. Appl. Meteorol. Climatol.* **47**, 1061–1076 (2008).
27. Rosenzweig, C. *et al.* Mitigating New York City's heat island: integrating stakeholder perspectives and scientific evaluation. *Bull. Am. Meteorol. Soc.* **90**, 1297–1312 (2009).
28. Hurrell, J. W. *et al.* The Community Earth System Model: a framework for collaborative research. *Bull. Am. Meteorol. Soc.* **94**, 1339–1360 (2013).
29. Oleson, K. *et al.* Technical Description of Version 4.0 of the Community Land Model (CLM) 257. Report No. NCAR/TN-478+STR (NCAR, 2010).
30. Qian, T. T., Dai, A., Trenberth, K. E. & Oleson, K. W. Simulation of global land surface conditions from 1948 to 2004. Part I: forcing data and evaluations. *J. Hydrometeorol.* **7**, 953–975 (2006).

Acknowledgements This research was supported by the Ministry of Education of China (grant PCSIRT), the Yale Climate and Energy Institute, the Yale Institute of Biospheric Studies, and a Yale University Graduate Fellowship. K.O. acknowledges support from NASA grant NNX10AK79G (the SIMMER project) and the NCAR WCIASP. NCAR is sponsored by the US National Science Foundation. The model simulations were supported by the Yale University Faculty of Arts and Sciences High Performance Computing Center.

Author Contributions X.L. designed the research. L.Z. carried out the model simulation and data analysis. R.B.S. contributed ideas to the research design. K.O. contributed ideas to the model simulation. X.L. and L.Z. drafted the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to X.L. (xuhui.lee@yale.edu).

Cooperating with the future

Oliver P. Hauser^{1,2*}, David G. Rand^{3,4*}, Alexander Peysakhovich^{1,3} & Martin A. Nowak^{1,2,5}

Overexploitation of renewable resources today has a high cost on the welfare of future generations^{1–5}. Unlike in other public goods games^{6–9}, however, future generations cannot reciprocate actions made today. What mechanisms can maintain cooperation with the future? To answer this question, we devise a new experimental paradigm, the ‘Intergenerational Goods Game’. A line-up of successive groups (generations) can each either extract a resource to exhaustion or leave something for the next group. Exhausting the resource maximizes the payoff for the present generation, but leaves all future generations empty-handed. Here we show that the resource is almost always destroyed if extraction decisions are made individually. This failure to cooperate with the future is driven primarily by a minority of individuals who extract far more than what is sustainable. In contrast, when extractions are democratically decided by vote, the resource is consistently sustained. Voting^{10–15} is effective for two reasons. First, it allows a majority of cooperators to restrain defectors. Second, it reassures conditional cooperators¹⁶ that their efforts are not futile. Voting, however, only promotes sustainability if it is binding for all involved. Our results have implications for policy interventions designed to sustain intergenerational public goods.

Providing for future generations is central to the survival of genes, families, organizations, nations and the global ecosystem^{1–5}. Yet providing for the future poses a challenge, as it requires making sacrifices today. Institutions can play an important role in promoting such cooperative behaviour among large groups of people. Traditionally, institutional designers have assumed that people are rational and purely self-interested, and proposed incentives that induce selfish people to cooperate^{17–19}.

In recent years, however, a large body of evidence has demonstrated that many people are not purely selfish^{5,20–25}. Here we consider the implications of these ‘social preferences’ for designing institutions that promote sustainability and intergenerational cooperation. We demonstrate that democracy can be a powerful institution for harnessing social preferences: although selfish people would vote for over-exploitation of resources, voting allows a pro-social majority to override a selfish minority (see Supplementary Information section 1 for further discussion).

To test this, we introduce a laboratory model of cooperating with the future—the Intergenerational Goods Game (IGG)—that builds on previous work using Public Goods Games^{7–9}, Common Pool Resource games^{6,11} and Threshold games^{4,26,27}. In these other games, selfishness creates social efficiency losses for the other members of one’s group. In contrast, the IGG is designed such that selfishness instead negatively affects subsequent groups.

In our IGG experiments, individuals form groups of five, which we refer to as generations. The first generation is endowed with a common pool of 100 units and each individual can extract between 0 and 20 units from the pool. If the total percentage of units extracted from the pool is at or below a commonly known extraction threshold, T , the pool will renew to 100 units for the next generation. If, however, the percentage extracted is above T , the pool is exhausted and all future generations receive no payoff (Fig. 1). After each generation, another generation occurs with probability δ , and with probability $1 - \delta$ the game ends: the discount factor δ models the extent to which the current generation values the next

generation (see Supplementary Information section 2 for further experimental details).

In the game theoretic tradition, the IGG framework is a great simplification relative to real-world intergenerational cooperation. For discussion of important aspects of intergenerational transfer which the IGG does not yet incorporate, as well as relation of our work to previous results on intergenerational transfer, see Supplementary Information section 3.

To explore behaviour in the IGG, we began with an ‘unregulated’ treatment: each group member individually chooses how many units to extract from the pool. We initialized 20 unregulated IGGs and passed each game’s pool across a series of generations with a discount factor of $\delta = 0.8$ (leading to an expected game length of five generations). For the pool to be replenished, each generation had to extract 50 units or less ($T = 50\%$). Thus, the socially efficient extraction (or ‘fair share’) was 10 units per individual on average. We focus on symmetric strategies and refer to individuals who extracted 10 or fewer units as cooperators, and those who extracted more than 10 units as defectors.

We found that a large majority of individuals cooperated (68%), in line with previous studies using non-student populations^{23–25}. Despite their good intentions, however, only 4 of the 18 games continuing on to a second generation had their pools sustained. These losses in sustainability compounded quickly over time: in the third generation, the number of refilled pools was down to two, and not a single refilled pool was available to the fourth generation (Fig. 2a). Notably, in most groups, only a minority of defectors was responsible for the exhaustion of the resource.

To address this sustainability failure, we introduce an institution that is firmly established in large parts of the world: democracy. Each group member votes for their generation’s extraction level, and the median vote is extracted by all players. Well studied by economists and political scientists^{10–15}, this ‘median voting’ rule guarantees socially optimal outcomes in a standard Public Goods Game, even with perfectly self-interested actors: the payoff-maximizing vote is full cooperation^{14,15}. In the IGG, however, this is not true: because the current group does not reap the benefits of cooperation, selfish players would vote to deplete the resource fully. From a traditional ‘public choice’ perspective based on rational self-interest, therefore, median voting is not attractive for promoting sustainability. If, however, enough players have social preferences, voting may be able to support sustainability in the IGG by allowing pro-social players to rein in selfish players. Thus a ‘behavioural public choice theorem’^{12–14} might favour median voting; see Supplementary Information section 1 for further discussion.

To explore the effects of median voting, we initialized another 20 IGGs using $\delta = 0.8$ and $T = 50\%$, and applied the voting rule. We found a dramatic increase in sustainability (Fig. 2b): all 20 common pools were sustained across all generations (unregulated versus voting: linear probability model (LPM) predicting pool sustainability at the generation level, $P < 0.001$; see Supplementary Information section 4 for statistical details).

Next we asked how robust the voting mechanism is to variation in the discount factor, δ , and the extraction threshold, T . In the experiments described above, there was an 80% chance that a future generation would exist ($\delta = 0.8$) and individuals had to sacrifice half of their possible payoff

¹Program for Evolutionary Dynamics, Harvard University, Cambridge, Massachusetts 02138, USA. ²Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ³Department of Psychology, Yale University, New Haven, Connecticut 06511, USA. ⁴Department of Economics, Yale University, New Haven, Connecticut 06511, USA. ⁵Department of Mathematics, Harvard University, Cambridge, Massachusetts 02138, USA.

*These authors contributed equally to this work.

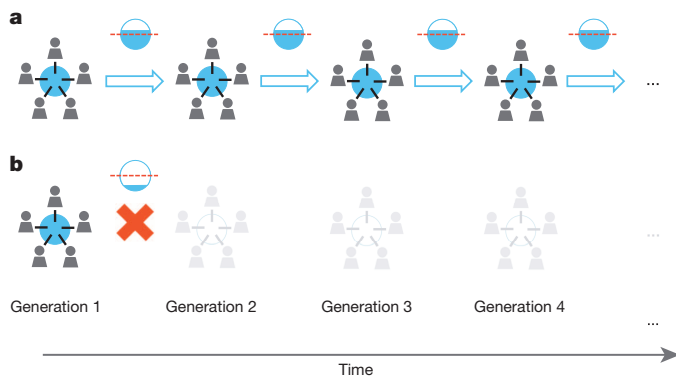


Figure 1 | An illustration of the Intergenerational Game (IGG). In each generation, a group of 5 people makes a decision (individually or according to an institutional rule) about their level of extraction from a common resource. **a**, If Generation 1's extractions do not violate the commonly known threshold, the resource refills and the same dilemma is presented to Generation 2. After each generation, another generation occurs with probability δ . **b**, If at any point the threshold requirement is not met, the resource does not renew and future generations receive no payoff. Maximal social welfare is achieved if no generation ever violates the threshold requirement by extracting too much from the common resource.

to extract a 'fair share' ($T = 50\%$). To assess robustness, we examined the effectiveness of voting in two treatments using lower δ values ($\delta = 0.7$ and $\delta = 0.6$, creating fewer future generations), and two other treatments using lower T values ($T = 40\%$ and $T = 30\%$, leading to a higher cost of cooperation). Each treatment again started with 20 pools.

We found that voting remained largely effective in promoting sustainability under these more adverse conditions (Fig. 2c). Although sustainability did vary significantly with δ (LPM, $P = 0.037$) and T (LPM, $P < 0.001$), the size of these effects was relatively small: decreasing δ or T by 0.1 decreases the probability of a pool being sustained by 4.6% or 14.6%, respectively. Moreover, under all conditions tested, voting led to much higher levels of sustainability than the original unregulated IGG (LPM, $P < 0.001$ for all comparisons).

The success of voting is driven by two factors. First, the decision-making power differs in the voting and unregulated institutions (Fig. 3a). In the voting institution, a majority of three cooperators who propose 10 unit extractions can overrule two defectors who propose 20 units. In contrast, if decisions are made at the individual level, a single defector can tip the balance of a group. In other words, voting allows a majority of cooperators to restrain a minority of defectors.

The second reason for the success of voting pertains to the psychology of social preferences. Median voting addresses the fears of players who care about future generations but worry that others (now or later) will exhaust the pool (that is, future-oriented 'conditional cooperators'¹⁶): as the outcome of the vote is applied to all players, everyone within a generation receives the same payoff and no one risks feeling like they have been taken advantage of. This, in turn, further increases the probability that a cooperative majority is formed and the pool is sustained, both in the current generation and in the future. Figure 3b is consistent with this assessment: the fraction of cooperators was 20% larger under voting than unregulated (LPM, $P < 0.001$).

Both of these factors predict that voting is only successful if everyone is bound by the outcome: a partial implementation¹⁵ provides an opportunity both for defectors to derail sustainability, and for potential cooperators to switch to defection out of fear that others will over-exploit.

We test this prediction by introducing a 'partial voting' treatment (another 20 pools, again using $\delta = 0.8$ and $T = 50\%$). Three of the five people in each generation are bound by the decision of a median vote among themselves. The other two people are not informed of the vote's outcome, and decide freely how much to extract. The sum of all five extractions is then compared to the extraction threshold T .

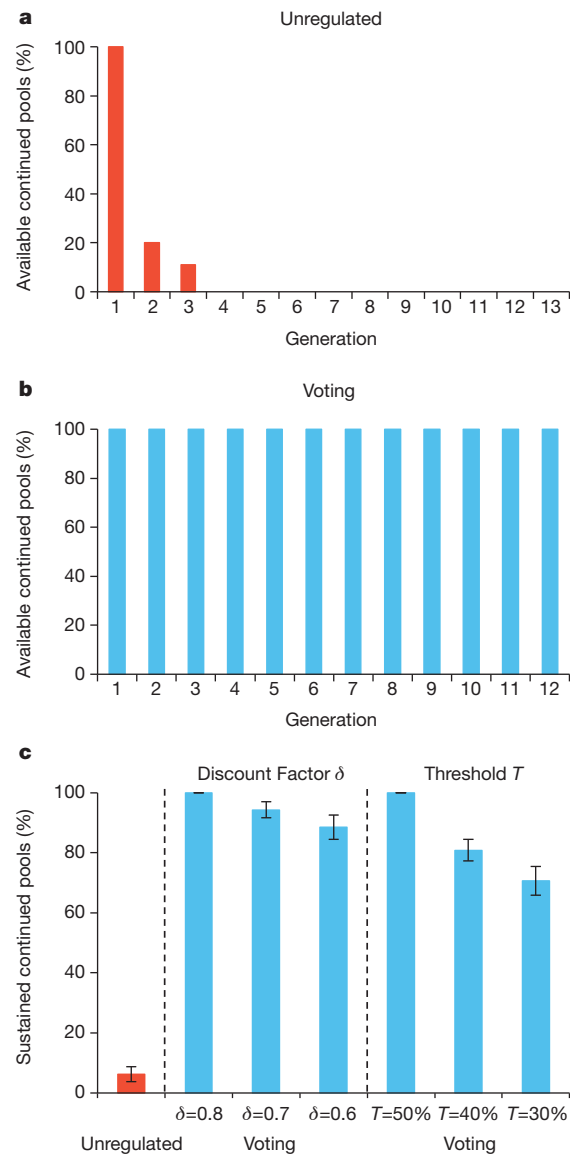


Figure 2 | Solving the (intergenerational) 'tragedy of the commons' through an institutional design. **a**, When decisions are made at the individual level, the availability of the common pools drastically decreases over time; $n = 480$. **b**, The introduction of a democratic voting institution strikingly improves sustainability; $n = 370$. **c**, Decreasing the discount factor from $\delta = 0.8$ to $\delta = 0.7$ ($n = 355$) or $\delta = 0.6$ ($n = 305$) while holding $T = 50\%$, or the extraction threshold from $T = 50\%$ to $T = 40\%$ ($n = 600$) or $T = 30\%$ ($n = 460$) while holding $\delta = 0.8$, increases the temptation to defect. Nonetheless, much less is extracted under median voting compared to the unregulated baseline. Error bars indicate standard errors of the mean.

As predicted, the partial voting institution was significantly less successful than the full voting institution (Fig. 4a, LPM, $P < 0.001$). This point was supported by bootstrapping simulations: of 10,000 pools created by randomly sampling participant decisions each generation, only 1.5% of available pools were sustained after 15 generations under partial voting, compared to 84% under full voting; see Extended Data Fig. 1 and Supplementary Information section 5 for details. We conclude that, for voting to effectively manage sustainability, it must be binding for all decision-makers.

In this paper, we have introduced a new laboratory model for cooperation across generations, the Intergenerational Goods Game (IGG). We have shown that in the absence of regulation, a minority of selfish players consistently deplete available resources. By implementing median voting, however, this negative outcome can be prevented—but only if all players are bound by the outcome of the vote. Votes that are only

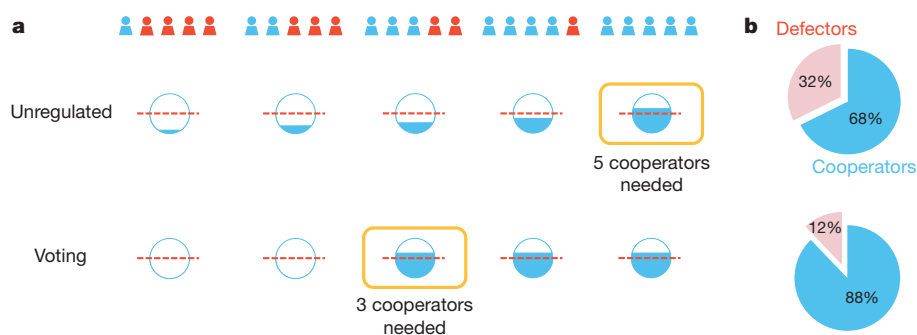


Figure 3 | The voting institution is robust to extreme decision-makers and thereby increases cooperative behaviour. **a**, The pivotal decision-maker in the voting institution is different from the unregulated institution. For instance, assume that $T = 50\%$ and that a cooperator and a defector always extract 10 and 20 units, respectively. The unregulated institution is vulnerable to extreme

decision-makers, whereas the voting institution is robust to a minority of defectors. This, in turn, bolsters the decision of those who are predisposed towards cooperation but fear to be exploited (for example, future-oriented 'conditional cooperators'). **b**, This leads to an increase of cooperators in the voting institution ($n = 370$) over the unregulated institution ($n = 480$).

partially binding, such as the international Kyoto protocol, have little power.

More generally, our results emphasize the importance of institutional designers moving away from the assumption of universal self-interest. We extend the 'behavioural public choice theorem'^{12–14} by demonstrating how voting can allow a majority of pro-social individuals to override a purely selfish minority, leading to costly group-level cooperation with future generations. Real-world data are consistent with this suggestion: countries that are more democratic also have more sustainable energy policies (combining data for 128 countries from *The Economist Democracy Index* and World Energy's Energy Sustainability Index, $P < 0.001$, $R^2 = 0.36$; robust to controlling for GDP, Gini index, population size, literacy rate, unemployment rate, life expectancy and level of corruption; see Extended Data Fig. 2 and Supplementary Information section 6 for details). Policy makers can do much to promote the public good by using a behavioural approach that is informed by a more accurate understanding of human psychology^{14,28–30}. Many citizens are ready to sacrifice for the greater good. We just need institutions that help them do so.

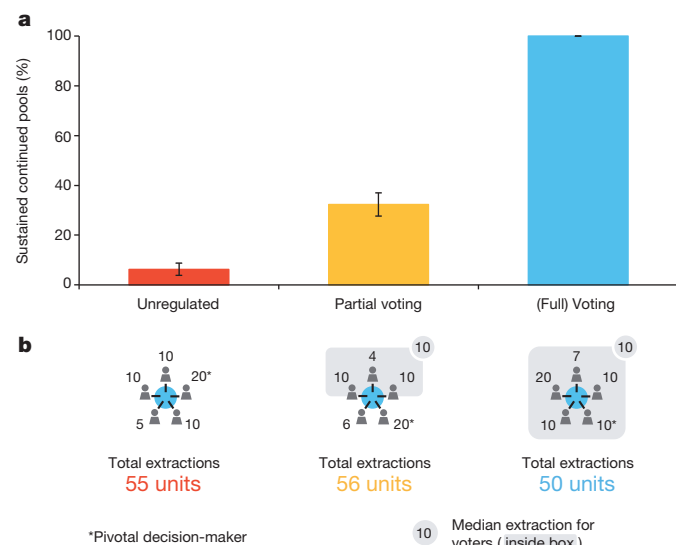


Figure 4 | Voting must be binding for all players in order to achieve high levels of sustainability. **a**, In a partially implemented voting institution ($n = 495$), three of the individuals are bound to vote while the other two can extract at will. A partially implemented voting institution is not robust to a minority of defectors and also cannot reassure conditional cooperators. Thus, partial voting fails to lead to sustainable outcomes. **b**, Three real sets of decisions from our data demonstrate a consequence of the pivotal extractor outside the voting group.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 May 2013; accepted 27 May 2014.

Published online 25 June 2014.

- Hardin, G. The tragedy of the commons. *Science* **162**, 1243–1248 (1968).
- Ostrom, E. *Governing the Commons: The Evolution of Institutions For Collective Action* (Cambridge Univ. Press, 1990).
- Levin, S. A. *Fragile Dominion: Complexity and the Commons* (Basic Books, 2000).
- Milinski, M., Semmann, D., Krambeck, H. J. & Marotzke, J. Stabilizing the Earth's climate is not a losing game: supporting evidence from public goods experiments. *Proc. Natl Acad. Sci. USA* **103**, 3994–3998 (2006).
- Wade-Benzoni, K. A. & Tost, L. P. The egoism and altruism of intergenerational behavior. *Pers. Soc. Psychol. Rev.* **13**, 165–193 (2009).
- Ostrom, E., Walker, J. & Gardner, R. Covenants with and without a sword: self-governance is possible. *Am. Polit. Sci. Rev.* **86**, 404–417 (1992).
- Milinski, M., Semmann, D., Bakker, T. C. M. & Krambeck, H. J. Cooperation through indirect reciprocity: image scoring or standing strategy? *Proc. R. Soc. Lond. B* **268**, 2495–2501 (2001).
- Fehr, E. & Gächter, S. Altruistic punishment in humans. *Nature* **415**, 137–140 (2002).
- Rand, D. G., Dreber, A., Ellingsen, T., Fudenberg, D. & Nowak, M. A. Positive interactions promote public cooperation. *Science* **325**, 1272–1275 (2009).
- Holcombe, R. G. The median voter model in public choice theory. *Public Choice* **61**, 115–125 (1989).
- Walker, J. M., Gardner, R., Herr, A. & Ostrom, E. Collective choice in the commons: experimental results on proposed allocation rules and votes. *Econ. J.* **110**, 212–234 (2000).
- Ertan, A., Page, T. & Putterman, L. Who to punish? Individual decisions and majority rule in mitigating the free rider problem. *Eur. Econ. Rev.* **53**, 495–511 (2009).
- Putterman, L., Tyran, J.-R. & Kamei, K. Public goods and voting on formal sanction schemes. *J. Public Econ.* **95**, 1213–1222 (2011).
- Kamei, K., Putterman, L. & Tyran, J.-R. State or nature? Formal vs. informal sanctioning in the voluntary provision of public goods. *Exp. Econ.* <http://dx.doi.org/10.1007/s10683-014-9405-0> (2014).
- Bernard, M., Dreber, A., Strimling, P. & Eriksson, K. The subgroup problem: When can binding voting on extractions from a common pool resource overcome the tragedy of the commons? *J. Econ. Behav. Organ.* **91**, 122–130 (2013).
- Fischbacher, U., Gächter, S. & Fehr, E. Are people conditionally cooperative? Evidence from a public goods experiment. *Econ. Lett.* **71**, 397–404 (2001).
- Coase, R. H. The problem of social cost. *J. Law Econ.* **3**, 1–44 (1960).
- Mueller, D. C. *Public Choice* (Cambridge Univ. Press, 1979).
- Williamson, O. E. *The Economic Institutions of Capitalism* (Simon & Schuster, 1985).
- Forsythe, R., Horowitz, J. L., Savin, N. E. & Sefton, M. Fairness in simple bargaining games. *Games Econ. Behav.* **6**, 347–369 (1994).
- Camerer, C. F. *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton Univ. Press, 2003).
- Charness, G. & Rabin, M. Understanding social preferences with simple tests. *Q. J. Econ.* **117**, 817–869 (2002).
- Fosgaard, T., Hansen, L. G. & Wengström, E. *Framing and misperceptions in a public good experiment*. Working paper 2011/11. (Institute for Food and Resource Economics, 2011).
- Amir, O., Rand, D. G. & Gal, Y. K. Economic games on the internet: the effect of \$1 stakes. *PLoS ONE* **7**, e31461 (2012).
- Rand, D. G., Greene, J. D. & Nowak, M. A. Spontaneous giving and calculated greed. *Nature* **489**, 427–430 (2012).
- Jacquet, J. et al. Intra- and intergenerational discounting in the climate game. *Nature Climate Change* **3**, 1025–1028 (2013).
- Cadsby, C. B. & Maynes, E. Gender and free riding in a threshold public goods game: experimental evidence. *J. Econ. Behav. Organ.* **34**, 603–620 (1998).

28. Oullier, O. Behavioural insights are vital to policy-making. *Nature* **501**, 463 (2013).
29. Benkler, Y. *The Penguin and the Leviathan: How Cooperation Triumphs Over Self-interest* (Random House, 2011).
30. Haynes, L., Service, O., Goldacre, B. & Torgerson, D. Test, learn, adapt: developing public policy with randomised controlled trials. UK Cabinet Office Behavioural Insights Team <http://dx.doi.org/10.2139/ssrn.2131581> (2012).

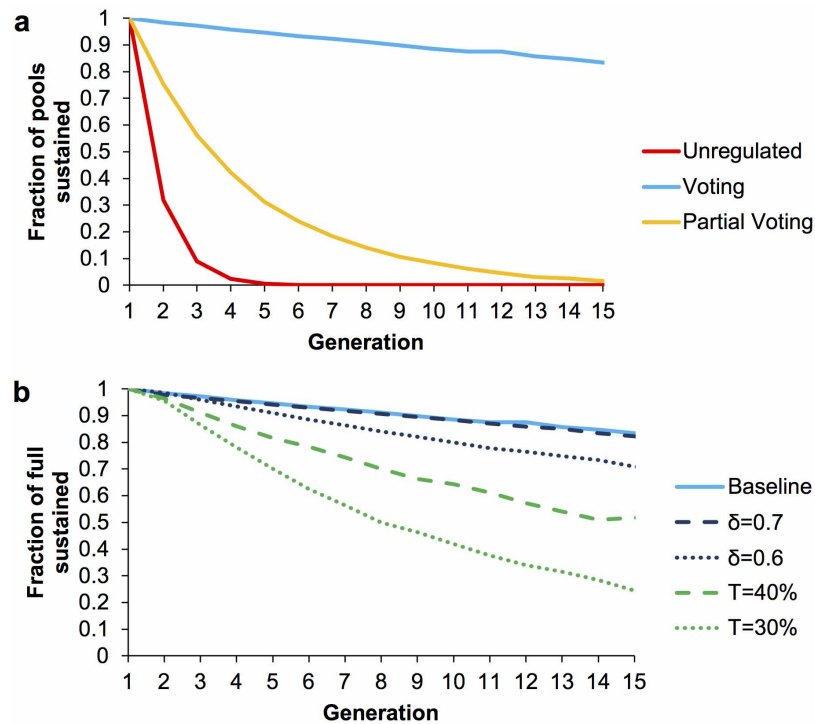
Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Dreber for discussion and three anonymous reviewers for helpful feedback. Financial support from the Department of Organismic

and Evolutionary Biology at Harvard, the Harvard Office for Sustainability and the John Templeton Foundation is gratefully acknowledged.

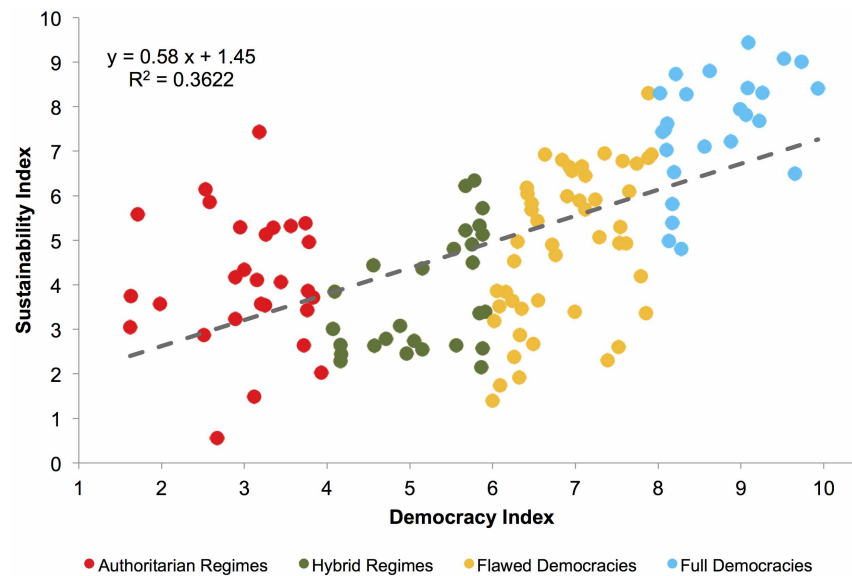
Author Contributions O.P.H., D.G.R., A.P. and M.A.N. designed and performed the experiments, analysed the data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.A.N. (martin_nowak@harvard.edu).



Extended Data Figure 1 | Bootstrapping simulations demonstrate the robustness of full voting and the failure of partial voting. We address sources of noise in the sequence of events that occurred in our experiment by conducting a set of computer simulations using the data generated by our participants. We randomly sample (with replacement) a series of generations of participant decisions, and calculate the fraction of those generations in which the pool was refilled. For each condition, we simulate 10,000 pools (or 1,000,000

pools if $\delta < 0.8$) for 15 generations. **a**, Simulated data for the unregulated, full voting and partial voting conditions show that full voting is by far the most successful at sustaining the pool. **b**, Simulated data for the $T = 40\%$, $T = 30\%$, $\delta = 0.7$ and $\delta = 0.6$ conditions shows that reducing δ has only a small effect, and although reducing T does undermine sustainability, the effect is much less striking than that of unregulated or partial voting despite the higher value of T in these less-regulated conditions.



Extended Data Figure 2 | Countries with more democratic governments have more sustainable energy policies. Energy sustainability index (as measured by the World Energy organization) is shown as a function of the democracy index (as measured by *The Economist* Intelligence Unit) for $n = 128$ countries. A strong positive association is clearly visible, and this association is robust to controlling for gross domestic product (GDP), Gini index, population

size, literacy rate, unemployment rate, life expectancy and level of corruption. Thus we provide preliminary empirical support for the role of democracy in promoting sustainability outside the laboratory. We adopt the colouring and naming scheme from *The Economist* Intelligence Unit's classification of regimes.

Amphibians acquire resistance to live and dead fungus overcoming fungal immunosuppression

Taegan A. McMahon^{1,2*}, Brittany F. Sears¹, Matthew D. Venesky³, Scott M. Bessler¹, Jenise M. Brown¹, Kaitlin Deutsch¹, Neal T. Halstead¹, Garrett Lentz¹, Nadia Tenouri¹, Suzanne Young¹, David J. Civitello¹, Nicole Ortega¹, J. Scott Fites⁴, Laura K. Reinert⁵, Louise A. Rollins-Smith^{4,5}, Thomas R. Raffel⁶ & Jason R. Rohr^{1*}

Emerging fungal pathogens pose a greater threat to biodiversity than any other parasitic group¹, causing declines of many taxa, including bats, corals, bees, snakes and amphibians^{1–4}. Currently, there is little evidence that wild animals can acquire resistance to these pathogens⁵. *Batrachochytrium dendrobatidis* is a pathogenic fungus implicated in the recent global decline of amphibians⁶. Here we demonstrate that three species of amphibians can acquire behavioural or immunological resistance to *B. dendrobatidis*. Frogs learned to avoid the fungus after just one *B. dendrobatidis* exposure and temperature-induced clearance. In subsequent experiments in which *B. dendrobatidis* avoidance was prevented, the number of previous exposures was a negative predictor of *B. dendrobatidis* burden on frogs and *B. dendrobatidis*-induced mortality, and was a positive predictor of lymphocyte abundance and proliferation. These results suggest that amphibians can acquire immunity to *B. dendrobatidis* that overcomes pathogen-induced immunosuppression^{7–9} and increases their survival. Importantly, exposure to dead fungus induced a similar magnitude of acquired resistance as exposure to live fungus. Exposure of frogs to *B. dendrobatidis* antigens might offer a practical way to protect pathogen-naïve amphibians and facilitate the reintroduction of amphibians to locations in the wild where *B. dendrobatidis* persists. Moreover, given the conserved nature of vertebrate immune responses to fungi⁵ and the fact that many animals are capable of learning to avoid natural enemies¹⁰, these results offer hope that other wild animal taxa threatened by invasive fungi might be rescued by management approaches based on herd immunity.

In recent decades, emerging fungal pathogens have had devastating effects on agriculture and caused population declines of several plant and animal species¹. Many plants can acquire immunological resistance to fungal pathogens, which has proven useful for managing fungal pathogens of crops¹¹, but acquired resistance to fungi in wild animals has not been well studied⁵. If natural variation in acquired resistance to fungi exists in wild animal populations, it might partly explain why fungal pathogens cause epidemics and extirpations of some animal populations (those that have not acquired resistance) but persist in an endemic state in others (those that have acquired resistance)^{12,13}. If wildlife managers could induce resistance in enough individuals of a population, they might be able to drive the basic reproductive ratio (R_0 ; the average number of infections one infected individual generates in a population of susceptible hosts over the course of its infectious period) of a pathogenic fungus below one, such that 'herd immunity' would protect even pathogen-naïve members of the population, a concept that is the basis for vaccination campaigns. Consequently, acquired resistance offers a potential tool to rescue animal populations threatened by fungi¹⁴.

Here, we investigate whether amphibians can acquire behavioural and immunological resistance to the fungal pathogen *Batrachochytrium dendrobatidis* (*Bd*). We define acquired resistance broadly as the process of reducing infection loads upon subsequent exposure to a pathogen or

antigen caused by changes in behaviour, innate immunity or adaptive immunity within the lifetime of an individual. We focused on *Bd* because it is implicated in the global decline of amphibians, the most threatened vertebrate taxon⁶. *Bd* is known to hinder amphibian lymphocyte responses⁷, but acquired resistance might still be detectable if its strength exceeds any immunosuppression by *Bd*. Hence, it is critical to quantify the net effect of any acquired host resistance and *Bd*-induced immunosuppression on *Bd* abundance (average number of *Bd* zoospores on *Bd*-exposed hosts) across multiple host species to evaluate the feasibility of acquired resistance as a conservation tool¹⁴. Although beneficial for understanding mechanisms of immunity, gene expression studies or studies that only quantify immunity to *Bd* (and not also *Bd* loads) do not capture the extended phenotype and thus cannot quantify the net effect of acquired immunity and immunosuppression on the host–parasite interaction^{8,9,15,16}. Surprisingly, there are no published studies that experimentally subjected amphibians to regimes of repeated *Bd* infection followed by clearance and then tested for an association between the number of previous infections and *Bd* abundance, immune parameters and behavioural avoidance (ref. 14, but see also 15–22). We hypothesized that, despite *Bd*-induced immunosuppression^{7–9}, effects of repeated exposures to *Bd* followed by pathogen clearance could include reduced *Bd* abundance on frog skin, increases in the abundance or efficacy of skin peptides, augmentation of the abundance of responding lymphocyte populations (mediators of *Bd* resistance^{16–18}), and induction of learned avoidance of *Bd*.

To create variation in number of exposures to *Bd*, groups of frogs were exposed to *Bd* and cleared of their infection using heat zero to four times (depending on their treatment assignment and experiment) in a manner that prevented a simple decay of immune responses from accounting for our results (Extended Data Table 1; Supplementary Methods). Frogs were swabbed for *Bd* before each experiment, after each *Bd* exposure, and after each clearance period. Quantitative PCR (qPCR) of these swabs revealed that frogs were free of *Bd* before each experiment, that all unexposed control frogs were *Bd*-free after each *Bd* growth period (that is, no cross contamination), that exposed frogs generally became infected (average prevalence across exposure periods was 85%), and that heat-clearances were 100% effective at eliminating established *Bd* infections (Extended Data Table 2). Additionally, our methods ensured that we did not confound acquired resistance, a form of phenotypic plasticity, with selection (via mortality of low-resistance individuals; see Supplementary Results, Extended Data Table 3, Extended Data Fig. 1) or *Bd* inoculates (via different inoculates for each treatment), confounding factors that have hampered previous studies¹⁴.

Many hosts are capable of avoiding pathogens, but few studies have tested whether avoidance is learned or innate^{23,24}. To determine if amphibians could learn to avoid *Bd*, we conducted two separate experiments to examine how number of exposures to *Bd* affected the amount of time that oak toads (*Bufo quercicus*) spent on a *Bd*-negative or *Bd*-inoculated side

¹University of South Florida, Department of Integrative Biology, Tampa, Florida 33620, USA. ²University of Tampa, Department of Biology, Tampa, Florida 33606, USA. ³Allegheny College, Department of Biology, Meadville, Pennsylvania 16335, USA. ⁴Vanderbilt University, Biological Sciences Department, Nashville, Tennessee 37232, USA. ⁵Vanderbilt University, School of Medicine, Departments of Pathology, Microbiology and Immunology and Pediatrics, Nashville, Tennessee 37232, USA. ⁶Oakland University, Department of Biology, Rochester, Michigan 48309, USA.

*These authors contributed equally to this work.

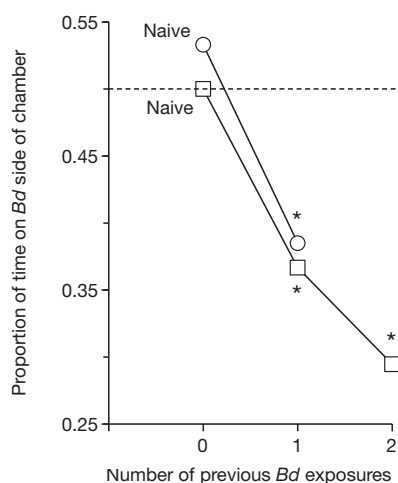


Figure 1 | Effects of 0, 1 or 2 previous exposures to live *Batrachochytrium dendrobatidis* (*Bd*) on the proportion of time that toads (*Bufo quercicus*) spent on the *Bd*-positive side of a test chamber in two experiments (Experiment 1: circles, Experiment 2: squares). In both experiments ($n = 30$ toads), *Bd*-naïve frogs showed no significant avoidance or attraction to *Bd* ($\chi^2_1 = 0.29$, $P = 0.59$), but frogs previously infected with *Bd* once or twice chose the *Bd*-free substrate more frequently than expected by chance ($\chi^2_1 = 6.7$, $P = 0.009$ and $\chi^2_1 = 9.7$, $P = 0.002$, respectively). Asterisks represent significant avoidance of *Bd*.

of a test chamber. *Bd*-naïve frogs showed no significant avoidance or attraction to *Bd* ($\chi^2_1 = 0.289$, $P = 0.591$), but frogs previously infected with *Bd* once or twice (all the exposures resulted in infections) chose the *Bd*-free substrate approximately 65% and 70% of the time, respectively (experiment 1: $\chi^2_1 = 6.683$, $P = 0.009$, experiment 2: $\chi^2_1 = 9.693$, $P = 0.002$; Fig. 1), resulting in a significant interaction between naivety and *Bd* avoidance (number of previous infections \times deviation from 50% null interaction: $\chi^2_1 = 9.107$, $P = 0.011$; Fig. 1). These results were consistent across the two experiments (that is, repeatable; Fig. 1). Importantly, this learned behavioural resistance should be unaffected by the immunosuppressive effects of *Bd*^{7–9}.

To test for acquired immunological resistance, we assessed whether the number of exposures to live *Bd* (0, 1, 2, 3 or 4 exposures) was a significant predictor of *Bd* abundance on Cuban treefrogs (*Osteopilus septentrionalis*) after the third and fourth *Bd* exposure and clearance regimens (Extended Data Table 1). Number of exposures to *Bd* was a significant negative predictor of *Bd* abundance on the frogs ($\chi^2_1 = 8.40$, $P = 0.003$; Fig. 2a and Extended Data Fig. 2a; see Extended Data Table 3 for prevalence data), with a 75% drop in *Bd* loads from the first to third *Bd* exposure (Fig. 2a, open circles). To evaluate whether the observed resistance to *Bd* could be attributable to changes in innate or cellular immunity, we quantified the abundance and efficacy of skin peptides and splenic leukocytes (enriched for lymphocytes) after only the fourth exposure and clearance (0–4 exposures) and then tested whether number of *Bd* exposures was a significant predictor of these responses. The observed acquired resistance to *Bd* did not seem to be strongly attributable to changes in skin peptides because the number of *Bd* exposures was not significantly correlated with skin peptide abundance ($\chi^2_1 = 0.7$, $P = 0.38$; Extended Data Fig. 3a) or efficacy at inhibiting *Bd* growth *in vitro* (Peptide treatment \times number of *Bd* exposures: $\chi^2_1 = 0.329$, $P = 0.566$; Extended Data Fig. 3b, Supporting Information). However, number of *Bd* exposures was a significant positive predictor of lymphocyte abundance in the spleen ($\chi^2_1 = 5.9$, $P = 0.015$, Fig. 3a) and lymphocyte proliferation in response to *Bd* ($\chi^2_1 = 9.5$, $P = 0.002$) and phytohaemagglutinin ($\chi^2_1 = 78.4$, $P < 0.0001$; number of exposures \times method of stimulation: $\chi^2_1 = 1.3$, $P = 0.24$; Fig. 3b). Phytohaemagglutinin is a mitogen that triggers T-lymphocyte cell division and thus functioned as a positive control for lymphocyte proliferation²⁵. Moreover, the slope of the relationship between *Bd* abundance and lymphocyte numbers in the spleen became more positive with each *Bd*

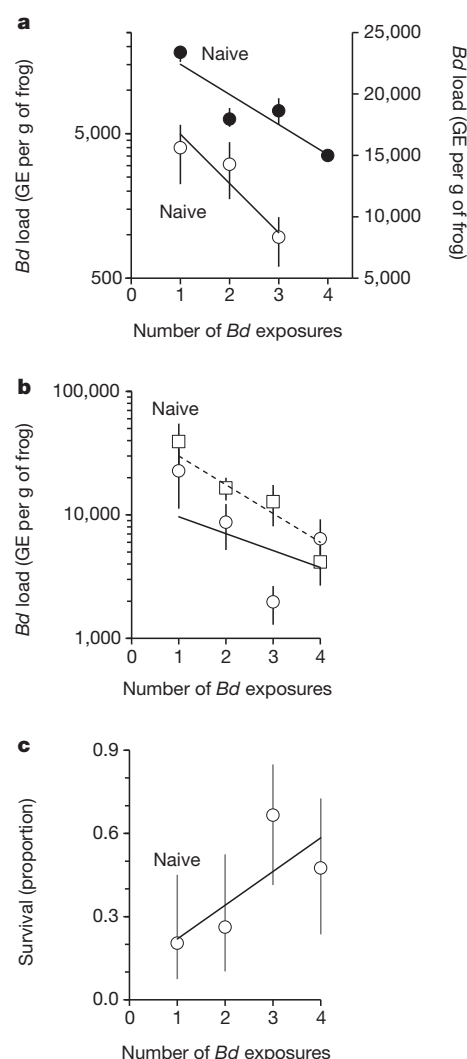


Figure 2 | Effects of 1–4 exposures to *Batrachochytrium dendrobatidis* (*Bd*) on *Bd* abundance per gram of frog and frog (*Osteopilus septentrionalis*) survival. **a**, **b**, Effects of exposures on mean *Bd* abundance (zoospore genome equivalents (GE)/g of frog) after exposure period 3 (\pm s.e.m., open circles and left axis; $\chi^2_1 = 8.40$, $P = 0.003$) and exposure period 4 (closed circles, right axis; bootstrapped means \pm bootstrapped 95% confidence interval, see Supplementary Methods for details) in the first immunological resistance experiment (**a**) and exposure period 4 in the second immunological resistance experiment (\pm s.e.m.; live *Bd* exposures (circles and solid line): $\chi^2_1 = 4.9$, $P = 0.02$; dead *Bd* exposures (squares and dotted line): $\chi^2_1 = 11.3$, $P < 0.001$) (**b**). **c**, Effects of live *Bd* exposures on mean frog survival (\pm 95% confidence interval; odds ratio: 1.66, $\chi^2_1 = 4.45$, $P = 0.035$; experiment treated as a temporal block). The best-fit lines are based on predicted values from the implemented zero-inflated negative binomial (*Bd* abundance) and binomial statistical models (frog survival). Naivety was based on the state of the frog before *Bd* exposure during the focal exposure period (third or fourth exposure period depending on what is being displayed). Thus, frogs exposed to *Bd* for the first time during the focal exposure period were classified as naïve because they had not previously been exposed to *Bd*.

exposure, indicating that lymphocytes were stimulated to proliferate with subsequent exposures (number of exposures \times lymphocyte abundance: $F_{1,54} = 4.4$, $P = 0.04$, Fig. 3c).

Two previous studies attempted to immunize frogs systemically through injections and did not find evidence of acquired protection against *Bd*^{21,26}, suggesting that the approach used in our study, exposure by way of the skin, might be critical for induction of acquired resistance. Consistent with this proposition are studies suggesting that the adaptive immune system of amphibians responds to *Bd*. For example, the heterozygosity

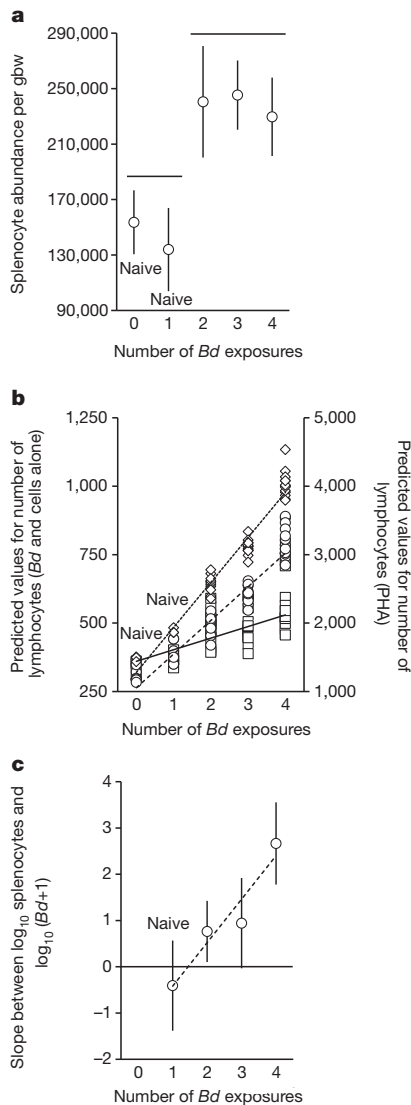


Figure 3 | Effects of 0–4 exposures to live *Batrachochytrium dendrobatidis* (*Bd*) on immune parameters of frogs (*Osteopilus septentrionalis*). **a**, Mean (\pm s.e.m.) lymphocyte abundance in the spleen (that is, splenocytes) per gram body weight (gbw). Splenocyte densities did not differ within the naive ($\chi^2_1 = 0.35$, $P = 0.56$) or experienced groups ($\chi^2_1 = 0.14$, $P = 0.93$) but differed significantly between these two groups ($\chi^2_1 = 11.35$, $P < 0.001$; designated by the horizontal lines). Frogs exposed to *Bd* for the first time during the fourth exposure period were classified as naive because they had not previously been exposed to *Bd*. **b**, Lymphocyte proliferation when cultured as splenocytes alone (squares and solid line), or with heat-killed *Bd* zoospores (circles and dashed line) or phytohaemagglutinin (PHA; diamonds and dotted line). Values provided are predicted values of proliferated lymphocytes (measured as counts per minute by a scintillation counter) based on a negative binomial model ($n = 19, 17, 17, 10$ and 7 , respectively for 0–4 previous live *Bd* exposures for both the splenocytes alone (that is, control) and splenocytes plus live *Bd* and $n = 15, 13, 11, 2$ and 4 , respectively, for splenocytes plus PHA). See text for statistics. **c**, Slope (\pm s.e.) of the relationship between *Bd* abundance and splenocyte abundance ($F_{1,54} = 4.4$, $P = 0.04$), showing greater lymphocyte proliferation in response to *Bd* load with each previous *Bd* exposure.

of genes associated with adaptive immunity (major histocompatibility complex loci) was greater in frog populations more resistant to *Bd*¹⁸, which would have been exposed cutaneously. Additionally, frogs irradiated to reduce their lymphocyte numbers had increased susceptibility to *Bd* infections, and frogs immunized against *Bd* via injection with heat-killed *Bd* cells had elevated levels of *Bd*-specific IgM and IgY serum antibodies¹⁷.

The only previous study to repeatedly infect and clear amphibians of *Bd* seemed to conflict with our results²⁰. In this study²⁰, the authors concluded that there was no evidence of acquired resistance in the critically endangered booroolong frog (*Litoria booroolongensis*) because frogs previously cleared of *Bd* with the fungicide itraconazole did not have significantly lower prevalence or mortality upon re-infection than frogs that were not previously exposed to *Bd* or the fungicide (20 of 32 versus 14 of 28 infected, respectively; $\chi^2_1 = 0.9$, $P = 0.33$). However, the fungicide alone increased *Bd* prevalence in *Bd*-naïve frogs (no previous fungicide: 14 of 28 infected, previous fungicide: 10 of 11 infected, $\chi^2_1 = 5.5$, $P = 0.01$), consistent with other studies suggesting that itraconazole is immunosuppressive²⁷. Hence, a better test for acquired resistance would have been to compare *Bd* prevalence of frogs previously infected and cleared of *Bd* by the fungicide to the prevalence of *Bd*-naïve frogs also previously exposed to the fungicide. A re-analysis of these data show that previous infections significantly reduced prevalence from 91% to 63% ($\chi^2_1 = 3.1$, $P = 0.03$) when making this more direct comparison. In summary, despite the immunosuppressive effects of *Bd*^{7–9}, the net effect of previous exposures to and clearances of *Bd*, across three studies and three species (*B. quercicus*, *O. septentrionalis*, *L. booroolongensis*), was to reduce *Bd* prevalence or abundance on frogs. Differences in the strength of this net effect (immunosuppression + acquired resistance) or variation in pathogenicity across *Bd* strains might explain host variation in susceptibility to *Bd*^{21,26}.

Next, we sought to test whether amphibians could acquire resistance to dead *Bd* because induction of acquired resistance through exposure to dead *Bd* might offer a practical management tool to protect amphibian populations¹⁴. We repeated our first immunological resistance experiment described above but added a dead *Bd* exposure treatment and tracked frog survival for 6 weeks after the last *Bd* inoculation to enhance our survival estimates. Number of exposures to live *Bd* in this second immunological resistance experiment was again a significant negative predictor of *Bd* abundance on *O. septentrionalis* ($\chi^2_1 = 4.9$, $P = 0.02$; Fig. 2b, Extended Data Fig. 2b), replicating our previous findings. Additionally, when blocking by experiment ($P < 0.001$), number of exposures to live *Bd* was a significant positive predictor of the probability of surviving the experiment (logistic regression: $\chi^2_1 = 4.4$, $P = 0.03$; Fig. 2c) and a positive predictor of time of death (Cox survival analysis: $\chi^2_1 = 4.6$, $P = 0.03$, parameter \pm s.e. = 0.224 ± 0.105 , hazard ratio = 0.799; Supplementary Fig. 3). Indeed, frogs with previous exposure to *Bd* were 5.57 (odds ratio) times more likely to survive until the end of the experiment than frogs that were naive to *Bd*.

Similar to our findings for exposure to live *Bd*, number of previous exposures to dead *Bd* was a significant negative predictor of *Bd* abundance on frogs ($\chi^2_1 = 11.3$, $P < 0.001$) and the magnitude of acquired resistance to dead and live *Bd* did not significantly differ ($\chi^2_1 = 0.99$, $P = 0.32$; Fig. 2b, Extended Data Fig. 2b). By the end of six weeks of *Bd* growth, frogs previously exposed to dead *Bd* three or four times lived longer than those exposed to dead *Bd* two times ($df = 2$, Wald = 7.22, $P = 0.027$), but overall, number of exposures to dead *Bd* was not a significant predictor of time of death ($\chi^2_1 = 0.02$, $P = 0.900$; Extended Data Fig. 4). Our power for detecting an effect of dead *Bd* on survival, however, was lower than it was for live *Bd* because we only conducted one rather than two experiments with dead *Bd*.

As a result of efforts by the IUCN Amphibian Ark network and other conservation initiatives, hundreds of threatened amphibian species have been removed from their *Bd*-positive habitats and are being bred in captivity^{14,28}. However, these amphibians often fail to re-establish when released at their sites of collection, presumably because of the persistence of *Bd* at these sites on tolerant hosts^{14,28,29}. Inducing acquired resistance in these captive-bred amphibians might allow for their successful re-establishment^{14,28}. Although additional research is necessary to quantify the efficacy of releasing dead *Bd* into water bodies to protect amphibians and the non-target effects of these releases, mathematical models suggest that, despite biotic and abiotic reservoirs for *Bd*, this strategy offers a promising management tool to reduce R_0 of *Bd* below one, which could be used to proactively prevent *Bd* epidemics and to rescue host populations already threatened by this pathogen^{12,13}. However, the efficacy

of this management option will depend on several factors, such as whether *Bd*-naïve larval amphibians can also acquire immunity, the role of biotic and abiotic reservoirs in maintaining *Bd*, and the extent and magnitude of variation in acquired resistance among amphibian species. Perhaps most importantly, given the conserved nature of the immune responses of vertebrates to fungi⁵ and that many animals are capable of learning to avoid natural enemies¹⁰, the results presented here offer hope that other wild animal taxa threatened by invasive fungi, such as bats, bees, and snakes^{1–4}, might be capable of acquiring resistance and might also be rescued by management approaches based on herd immunity.

METHODS SUMMARY

To infect frogs, we pipetted either live *Bd* zoospores or control water lacking zoospores directly onto the ventral surface of each frog to prevent *Bd* avoidance. All frogs were held at 17 °C for 11 days to allow for *Bd* growth and then swabbed 10 times from hip to toe for *Bd* quantification by qPCR. Following the infection period, all frogs were moved to 23 °C for 4 days, to 30 °C for 11 days (see Methods for exception), and swabbed again to ensure that any infection was cleared. This pattern was reversed to return frogs to 17 °C. This procedure was repeated so that we had groups of frogs exposed to and cleared of *Bd* 0–4 times depending on the experiment (Extended Data Tables 1, 3). In the second immunological resistance experiment, groups of frogs were exposed to dead *Bd* every 2 days (because *Bd* DNA was undetectable after 2 days) for 11 days to match the continuous exposure to live *Bd* in this same experiment. These frogs were then challenged with live *Bd* after the last 11-day exposure to dead *Bd*. We followed the qPCR procedures described previously³⁰ to quantify *Bd* abundance from swabs. In the first immunological resistance experiment, we used methods published previously^{25,26} to quantify the abundance and efficacy of frog skin peptides and splenic lymphocytes.

Frog mass was used as a covariate in statistical analyses when significant. We treated number of *Bd* exposures as a continuous predictor in generalized linear models with the following response variables: frog growth, skin peptide abundance (normal errors), lymphocyte abundance (negative binomial errors), *Bd* abundance (zero-inflated negative binomial errors), *Bd* prevalence, and frog survival (binomial errors). We used generalized linear mixed effects models to determine if number of *Bd* exposures affected skin peptide efficacy, lymphocyte proliferation in response to *Bd* and phytohaemagglutinin (random effect: individual frog; normal and negative binomial errors, respectively) and whether toads significantly avoided *Bd* (random effects: frog and experiment; binomial errors).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 February; accepted 15 May 2014.

1. Fisher, M. C. *et al.* Emerging fungal threats to animal, plant and ecosystem health. *Nature* **484**, 186–194 (2012).
2. Blehert, D. S. *et al.* Bat white-nose syndrome: an emerging fungal pathogen? *Science* **323**, 227 (2009).
3. Cameron, S. A. *et al.* Patterns of widespread decline in North American bumble bees. *Proc. Natl Acad. Sci. USA* **108**, 662–667 (2011).
4. Allender, M. C. *et al.* *Chrysosporium* sp. infection in eastern massasauga rattlesnakes. *Emerg. Infect. Dis.* **17**, 2383–2384 (2011).
5. Sexton, A. C. & Howlett, B. J. Parallels in fungal pathogenesis on plant and animal hosts. *Eukaryot. Cell* **5**, 1941–1949 (2006).
6. Stuart, S. N. *et al.* Status and trends of amphibian declines and extinctions worldwide. *Science* **306**, 1783–1786 (2004).
7. Fites, J. S. *et al.* The invasive chytrid fungus of amphibians paralyzes lymphocyte responses. *Science* **342**, 366–369 (2013).
8. Rosenblum, E. B. *et al.* Genome-wide transcriptional response of *Silurana* (*Xenopus*) *tropicalis* to infection with the deadly chytrid fungus. *PLoS ONE* **4**, e6494 (2009).
9. Ribas, L. *et al.* Expression profiling the temperature-dependent amphibian response to infection by *Batrachochytrium dendrobatidis*. *PLoS ONE* **4**, e8408 (2009).
10. Chivers, D. P. & Smith, R. J. F. Chemical alarm signalling in aquatic predator-prey systems: A review and prospectus. *Ecoscience* **5**, 338–352 (1998).
11. Durrant, W. E. & Dong, X. Systemic acquired resistance. *Annu. Rev. Phytopathol.* **42**, 185–209 (2004).
12. Woodhams, D. C. *et al.* Mitigating amphibian disease: strategies to maintain wild populations and control chytridiomycosis. *Front. Zool.* **8**, 8 (2011).
13. Briggs, C. J., Knapp, R. A. & Vredenburg, V. T. Enzootic and epizootic dynamics of the chytrid fungal pathogen of amphibians. *Proc. Natl Acad. Sci. USA* **107**, 9695–9700 (2010).

14. Venesky, M. D., Raffel, T. R., McMahon, T. A. & Rohr, J. R. Confronting inconsistencies in the amphibian-chytridiomycosis system: implications for disease management. *Biol. Rev. Camb. Philos. Soc.* **89**, 477–483 (2014).
15. Richmond, J. Q., Savage, A. E., Zamudio, K. R. & Rosenblum, E. B. Toward immunogenetic studies of amphibian chytridiomycosis: linking innate and acquired immunity. *Bioscience* **59**, 311–320 (2009).
16. Rollins-Smith, L. A., Ramsey, J. P., Pask, J. D., Reinert, L. K. & Woodhams, D. C. Amphibian immune defenses against chytridiomycosis: impacts of changing environments. *Integr. Comp. Biol.* **51**, 552–562 (2011).
17. Ramsey, J. P., Reinert, L. K., Harper, L. K., Woodhams, D. C. & Rollins-Smith, L. A. Immune defenses against *Batrachochytrium dendrobatidis*, a fungus linked to global amphibian declines, in the south african clawed frog, *Xenopus laevis*. *Infect. Immun.* **78**, 3981–3992 (2010).
18. Savage, A. E. & Zamudio, K. R. MHC genotypes associate with resistance to a frog-killing fungus. *Proc. Natl Acad. Sci. USA* **108**, 16705–16710 (2011).
19. Murphy, P. J., St-Hilaire, S. & Corn, P. S. Temperature, hydric environment, and prior pathogen exposure alter the experimental severity of chytridiomycosis in boreal toads. *Dis. Aquat. Organ.* **95**, 31–42 (2011).
20. Cashins, S. D. *et al.* Prior infection does not improve survival against the amphibian disease chytridiomycosis. *PLoS ONE* **8**, e56747 (2013).
21. Stice, M. J. & Briggs, C. J. Immunization is ineffective at preventing infection and mortality due to the amphibian chytrid fungus *Batrachochytrium dendrobatidis*. *J. Wildl. Dis.* **46**, 70–77 (2010).
22. Shaw, S. D. *et al.* Experimental infection of self-cured *Leiopelma archeyi* with the amphibian chytrid *Batrachochytrium dendrobatidis*. *Dis. Aquat. Organ.* **92**, 159–163 (2010).
23. Rohr, J. R., Swan, A., Raffel, T. R. & Hudson, P. J. Parasites, info-disruption, and the ecology of fear. *Oecologia* **159**, 447–454 (2009).
24. Kiesecker, J. M., Skelly, D. K., Beard, K. H. & Preisser, E. Behavioral reduction of infection risk. *Proc. Natl Acad. Sci. USA* **96**, 9165–9168 (1999).
25. Rollins-Smith, L. A., Parsons, S. C. V. & Cohen, N. During frog ontogeny, PHA and Con-A responsiveness of splenocytes precedes that of thymocytes. *Immunology* **52**, 491–500 (1984).
26. Rollins-Smith, L. A. *et al.* Immune defenses of *Xenopus laevis* against *Batrachochytrium dendrobatidis*. *Front. Biosci.* **51**, 68–91 (2009).
27. Pawelec, G., Ehninger, G., Rehbein, A., Schaudt, K. & Jaschonek, K. Comparison of the immunosuppressive activities of the antimycotic agents, itraconazole, fluconazole, ketoconazole and miconazole on human T-cells. *Int. J. Immunopharmacol.* **13**, 299–304 (1991).
28. Venesky, M. D., Mendelson, J. R., Stiling, P., Sears, B. F. & Rohr, J. R. Selecting for tolerance against pathogens and herbivores to enhance success of reintroduction and translocation. *Conserv. Biol.* **26**, 586–592 (2012).
29. McMahon, T. A. *et al.* Chytrid fungus *Batrachochytrium dendrobatidis* has nonamphibian hosts and releases chemicals that cause pathology in the absence of infection. *Proc. Natl Acad. Sci. USA* **110**, 210–215 (2013).
30. Hyatt, A. D. *et al.* Diagnostic assays and sampling protocols for the detection of *Batrachochytrium dendrobatidis*. *Dis. Aquat. Organ.* **73**, 175–192 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank W. Holden and J. Pask for their assistance with the immunological assays, J. Cohen for comments on the manuscript, and our undergraduate assistants H. Folse, S. Lopez, L. Soto, S. Hekkanen and E. Creasey. We also thank V. Vasquez and J. Longcore for providing the *Bd* isolates used in these experiments. Funds were provided by grants from the National Science Foundation (DEB 0516227 and EF-1241889 to J.R.R. and IOS-1121758 to L.A.R.-S.), the National Institutes of Health (R01GM109499 to J.R.R.), the US Department of Agriculture (NRI 2006-01370 and 2009-35102-0543 to J.R.R.), the US Environmental Protection Agency grant (STAR R83-3835 and CAREER 83518801 to J.R.R.), and the NSF RCN “Refining and Diversifying Ecoimmunology”. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions T.A.M., T.R.R., J.R.R., B.F.S., N.T.H. and J.M.B. conceived and designed the first immunological resistance experiment, T.A.M., J.R.R., B.F.S., S.M.B., N.T.H., N.O. and J.M.B. conceived and designed the second immunological resistance experiment, M.D.V. and J.R.R. conceived and designed the behavioural resistance experiment. T.A.M. directed the first and second immunological resistance experiments and M.D.V. directed the behavioural resistance experiments. T.A.M., B.F.S., S.M.B., N.T.H., J.M.B., G.L., N.T., S.Y. and M.D.V. conducted the first and second immunological resistance experiments; M.D.V. and K.D. conducted the behavioural resistance experiments, and T.A.M., L.K.R., J.S.F. and L.A.R.-S. conducted the antimicrobial peptide collection and lymphocyte assays. T.A.M. and J.R.R. conducted the statistical analyses for the first and second immunological resistance experiments and D.J.C. consulted on these analyses; M.D.V. and J.R.R. conducted the statistical analyses for the behavioural resistance experiment. J.R.R. wrote the manuscript and handled all submissions and revisions. T.A.M. and B.F.S. wrote parts of the methods, and all authors contributed to its editing.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.A.M. (taeganmcmahon@gmail.com) or J.R.R. (jasonrohr@gmail.com).

A dynamic model of bovine tuberculosis spread and control in Great Britain

Ellen Brooks-Pollock^{1,2}, Gareth O. Roberts³ & Matt J. Keeling²

Bovine tuberculosis (TB) is one of the most complex, persistent and controversial problems facing the British cattle industry, costing the country an estimated £100 million per year¹. The low sensitivity of the standard diagnostic test leads to considerable ambiguity in determining the main transmission routes of infection, which exacerbates the continuing scientific debate^{2–6}. In turn this uncertainty fuels the fierce public and political disputes on the necessity of controlling badgers to limit the spread of infection. Here we present a dynamic stochastic spatial model for bovine TB in Great Britain that combines within-farm and between-farm transmission. At the farm scale the model incorporates stochastic transmission of infection, maintenance of infection in the environment and a testing protocol that mimics historical government policy. Between-farm transmission has a short-range environmental component and is explicitly driven by movements of individual cattle between farms, as recorded in the Cattle Tracing System². The resultant model replicates the observed annual increase of infection over time as well as the spread of infection into new areas. Given that our model is mechanistic, it can ascribe transmission pathways to each new case; the majority of newly detected cases involve several transmission routes with moving infected cattle, reinfection from an environmental reservoir and poor sensitivity of the diagnostic test all having substantive roles. This underpins our findings on the implications of control measures. Very few of the control options tested have the potential to reverse the observed annual increase, with only intensive strategies such as whole-herd culling or additional national testing proving highly effective, whereas controls focused on a single transmission route are unlikely to be highly effective.

Bovine tuberculosis (TB) has been a long-standing problem in Britain. In the first half of the 20th century, 40% of British cattle were suspected to be infected, with contaminated milk being a major transmission route to humans³. This route was eliminated with the pasteurization of milk, and a 'test-and-slaughter' scheme was started in the 1950s to eradicate infection among cattle. Incidence fell markedly and in the 1970s only 0.22% of tests revealed infected animals^{4,5}. However, over the past 20 years incidence has been steadily increasing, most notably in the south-west and west of England and the southwest of Wales. The 2001 foot-and-mouth disease (FMD) epidemic exacerbated the problem, as a reduction in bovine TB testing was followed by atypical movements of cattle to replenish FMD-affected farms⁶. Since then, results of the randomized badger culling trial have been reported and analysed⁷ and a variety of new legislation and controls have been introduced⁸.

Present-day control measures have evolved from the original test-and-slaughter scheme, but are still based on the regular testing of herds, slaughter of positive animals and the imposition of movement restrictions following a failed test^{5,8}. The most common method for detecting infected animals is the single intradermal comparative cervical tuberculin (SICCT) test (Extended Data Fig. 1), based on the measurement of the reactions produced to an antigen mixture (tuberculin) derived from *Mycobacterium bovis* and *Mycobacterium avium*⁹. The test sensitivity (proportion of infected animals successfully detected) is estimated to lie in the range 70–90%^{10–12}; in addition, an animal is unlikely to test

positive for a period of time following infection. Using such a test necessitates the use of multiple follow-up tests following a positive test on a herd to increase the chance of detecting all infected cattle. Movement restrictions are placed on herds until they pass one or two follow-up tests at approximately 60-day intervals, and once movement restrictions are lifted two further tests are required after 6 and 12 months. In 2010–2012 there was a marked expansion and consolidation of testing protocols in England, culminating in 2013 with the division of the country into annual testing counties in the south and west and four yearly testing counties in the east and north⁸. All herds in Wales were placed on annual testing in 2009. In 2013 over 8 million cattle were tested and over 32,000 cattle slaughtered to control this infection.

Bovine TB incidence can be measured in several different ways according to the results of the bovine TB testing policy. The number of reactors is defined as the total number of cattle testing positive. Failed tests are a herd-level measure and refer to any test on a farm (including routine, whole-herd and follow-up test) where at least one animal tests positive. Herd breakdowns (or Officially TB Free Withdrawn or Suspended, OTFW/S) are defined as a failed test on a herd not currently subjected to movement restrictions.

Mathematical modelling has formed an integral part in planning the control of livestock diseases such as FMD^{13,14}. Dynamic mechanistic modelling of bovine TB has posed a challenge owing to the long time-scales associated with the disease, the asymptomatic nature of infection, the ambiguity in the transmission pathways, the potential contribution from a local reservoir and the effect of complex and changing control policies^{12,15,16}. However, the availability of data from the Cattle Tracing System has enabled detailed investigation of livestock infections that are spread by the movement of cattle^{17,18}. Here we develop a mechanistic model of bovine TB transmission between the approximately 134,000 livestock premises in Great Britain that moved cattle between 1996 and 2011. Mechanistic models can be challenging to fit to observed data, as they are constrained by mechanistic transmission rules; however, they allow us to separate cause and effect and to determine the long-term consequences of alternative interventions.

A stochastic SEI (Susceptible, Exposed (or Latent), Infectious) process, the most relevant model formulation for tuberculosis¹⁹, is used to capture the infection dynamics within each farm. Within-farm transmission is modelled as both direct cattle-to-cattle and indirect from the farm environment, which increases persistence of infection. This environment captures the effect of both contaminated pasture and infected wildlife, with the two being inseparable in our formulation. Spread between farms is captured by the daily movement of approximately 30,000 cattle that can potentially transport infection, and from the local regional environment that facilitates both short distance spread and persistence. This regional environment again corresponds to both contaminated pasture and wildlife, and allows transmission between farms within the same parish. A herd-level testing policy is incorporated that explicitly uses recorded routine test dates and generates follow-up tests in accordance with historical government policy. As the model is a herd-level model that does not distinguish between individual cattle on a farm, we included all herd-level tests but no individual cattle tests.

¹Disease Dynamics Unit, Department of Veterinary Medicine, University of Cambridge, Madingley Road, Cambridge CB3 0ES, UK. ²WIDER Centre, Mathematics Institute and School of Life Sciences, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK. ³Department of Statistics, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK.

This model unifies within- and between-herd transmission; it is this and the combination of several key features that distinguishes this model from previous work^{12,18,20–22}. First, it is dynamic rather than statistical, such that the behaviour one day is carried forward and affects future events. Such an approach is vital in predicting the effects of alternative control measures rather than simply associating risks. Second, it deals with the infection dynamics at several spatial scales, incorporating within-farm behaviour into a national model of bovine TB spread. This level of detail is necessary, given that the transmission of infection within farms is slow and often only a few cattle are infected. In addition, the movement of cattle is modelled explicitly, such that susceptible, latent and infectious animals can all be moved on or off a farm, in contrast to traditional network-based models of disease transmission. The model includes a testing regime that approximately captures changing government policy. Throughout we make the important distinction between infected herds and herds where infection has been detected, which enables the investigation of realistic control measures.

The model is parameterized using sequential Monte Carlo approximate Bayesian computation^{23,24} (SMC-ABC), in which the posterior distributions of parameters are generated by repeated stochastic simulation of the model (see Methods, Supplementary Information, Extended Data Fig. 2 and Extended Data Table 1). We fit the model to the number of reactors and number of failed herd tests per county per year between 1997 and 2007. ABC offers a flexible method for inferring parameters, as standard likelihood-based approaches are infeasible because of the model complexity and the uncertainty of test results. Some key implications arise from this parameter inference. First, the test sensitivity is estimated to be 72% (credible interval 63–81%). This indicates that farms with a low number of infected animals may escape detection; at least four animals on a farm need to be infectious and tested to correctly identify a breakdown with 99% certainty. In addition, even when the farm is correctly detected individual animals may escape; on a farm with four infectious cattle there is only a 26% (12–49%) chance of detecting all four animals at the first test. Second, our inferred parameters indicate that the majority of within-farm transmission is through the environment rather than direct transmission (Extended Data Table 1), where this environment captures the effect of both contaminated pasture or infected wildlife. These localized environmental reservoirs also allow infection to persist between tests after the infectious animals have been removed. The level of infection in the environment is predicted to decay with a half-life of 34 days (credible interval 20–71 days) (Extended Data Table 1), and although this is rapid it is sufficient to allow infection to arise months after infectious cattle are removed.

The model, initialized to capture the detected cases in 1996–1998, is iterated forwards and captures many aspects of the spatio-temporal spread of bovine TB (Fig. 1, Extended Data Figs 3, 4). Although the fitting procedure operates at the county scale there is good agreement with the national observations, capturing steady rise in the annual number of reactors (cattle failing the test) (Fig. 1a), which have increased by on average 10% each year between 2004 and 2010. At the county scale we find an extremely strong correlations (Pearson correlation coefficient = 0.993, $P < 2 \times 10^{-16}$) between predictions and observations for both numbers of failed tests and numbers of reactors (Extended Data Fig. 3). In particular, we can visually compare the predicted and observed spatial spread of infection as it has expanded and intensified from 1998 to 2010 (Fig. 1b, Extended Data Fig. 4). Additionally, predictive models such as this can be used to extrapolate future behaviour, we do this here by reusing the historical cattle movements from 2007–2010 to predict infection up to 2015. Although individual movements vary from year to year, the overall pattern is largely consistent²⁰, suggesting that this replaying of annual patterns is a reasonable approximation. In the absence of additional controls, the recent trend is predicted to continue, with numbers of reactors continuing to grow exponentially; this increase tends to be concentrated within South Western counties with limited spatial increase in the five years from 2010 to 2015.

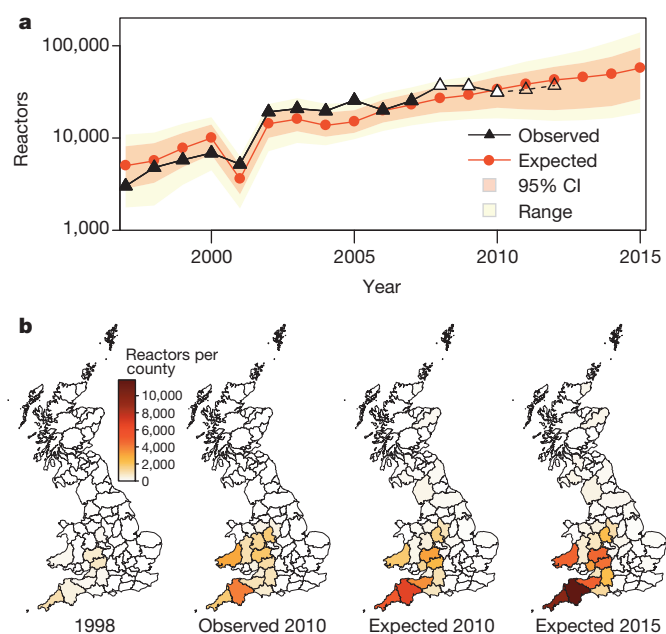


Figure 1 | Spatio-temporal comparison of model and data. **a**, Total number of reactors (cattle failing a test) per year. Triangles show the recorded data (filled black, years used for inference; filled white, years when detailed records are available; empty, years where only aggregate data are available). Red dots are the mean of 5,000 simulations; red shaded region is the combined 95% credible interval, whereas the yellow region is the range from 5,000 simulations. **b**, Spatial distribution of reactors at the county scale—showing observations from 1998 (initial conditions) and 2010; and mean of model replicates for 2010 and 2015.

We now use the model to investigate the driving mechanisms of transmission (Fig. 2). The farm-to-farm reproductive number measures the number of new secondary farms infected by each farm during its time harbouring infected animals (Fig. 2a). Transmission between farms occurs by two basic routes: either through the movement of infected cattle, or infection spread via the regional environment; we present these two separately. Movement is responsible for the majority (84%) of newly infected farms, with an over-dispersed distribution—where some farms are predicted to generate many secondary cases whereas around 90% of farms do not generate any secondary cases while harbouring infection. A similar pattern is predicted for transmission through the regional environment. This suggests that a small minority of farms act as superspreaders of infection, and therefore represents a more extreme version of the 80–20 rule (where 20% of farms are responsible for 80% of infection) reported for infection spread by cattle movements¹⁷.

From our model results, we are also able to identify the three contributing causes of each herd breakdown (Fig. 2b): (1) movement of an infected animal onto the farm has occurred since the last clear test (red), (2) transmission of infection from the environment (including both the local farm and regional environment) has occurred since the last clear test (green), (3) where the last clear test missed (failed to detect) infected animals that have remained on the farm (blue), and combinations of these three. We predict that no single cause dominates: 40% of herd breakdowns have infections due to all three causes, and only 29% of herd breakdowns are attributable a single cause. Although transmission from the environment is the single largest contributing factor, even removing this route completely would only eliminate 15% of breakdowns. This multifactorial nature of transmission contributes to the difficulties of controlling bovine TB.

We now use the model to go beyond a retrospective analysis of the past, to consider the effect of alternative control mechanisms, as though they had been implemented in 2005 and run for six years until the end of 2010 (Fig. 3, Extended Data Table 2, Extended Data Fig. 5). We investigated

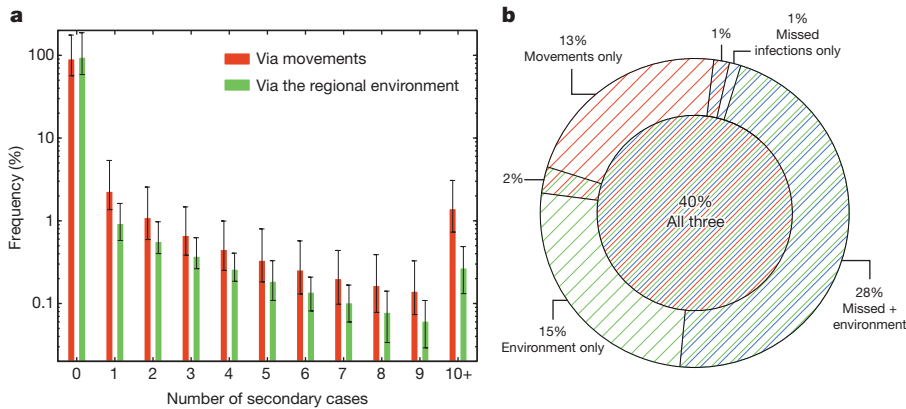


Figure 2 | Mechanisms driving transmission. **a**, Frequency distribution for infected farms generating a particular number of secondary infected farms (red, generated through movements; green, through local regional environment); note frequency is plotted on a logarithmic scale. **b**, Predicted causes of herd breakdowns (HBDs) in 2010. The area of each section corresponds to the proportion attributable to each cause (red, movements; green, local regional environment; blue, infected animals remain on a farm after testing). When multiple causes exist the area is shaded appropriately.

several control strategies (see Supplementary Material, Extended Data Table 2), but highlight here only those that have a substantial effect on the patterns of infection. Given that the cost of any control strategy is multifactorial, we show four distinct measures: the number of reactors, the number of cattle slaughtered, the number of herds tested and the number of herds under movement restrictions. Any strategy that can simultaneously lower all four of these is clearly beneficial.

Culling the entire herd if an animal tests positive (red bar) has by far the greatest effect, reducing the numbers of reactors, cattle slaughtered and farms under movement restrictions by over 80% compared to baseline in 2010 after 6 years of implementation. However, this strategy would necessitate a huge increase (approximately 20 fold) in the number of cattle slaughtered in the first year of operation; although this might be an acceptable cost if one is prepared to take a sufficiently long-term view. Vaccination of cattle which acts to slow the progression of infection (yellow), has a marked effect on all measures of bovine TB, but given

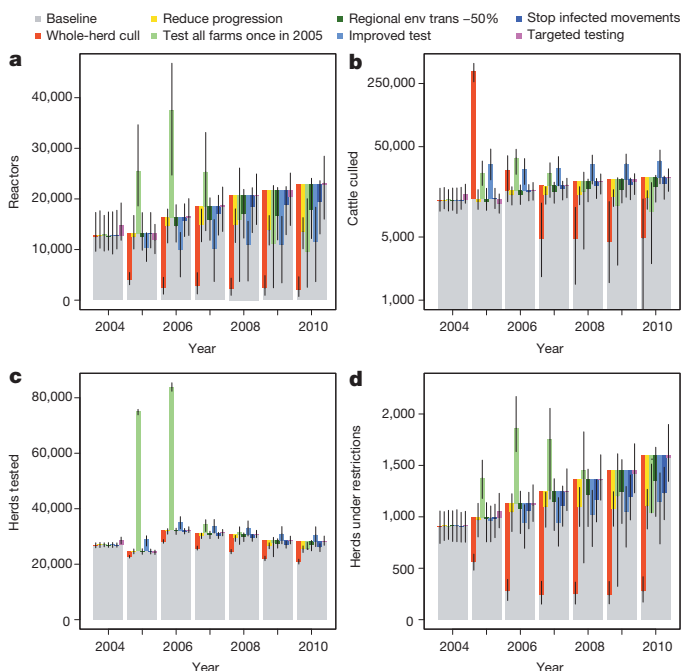


Figure 3 | Effect of different interventions assumed to begin in 2005 compared to a baseline of standard testing. **a–d**, Number of reactors in each year (**a**), number of slaughtered cattle (plotted on a logarithmic scale) (**b**), number of herds tested (**c**), number of herds under movement restrictions (**d**). Coloured bars show the mean deviation from the baseline scenarios; large grey bars show the baseline. The error bars represent the 95% combined prediction and credible interval from 1,000 simulations.

that the vaccine offers only limited protection it does not have the effect that we usually associate with mass-vaccination campaigns; in particular even assuming 100% uptake and ignoring interference with the SICTT, vaccination is just sufficient to overturn the observed annual increase. A single additional mass testing program in 2005 (pale green), comparable to what was performed in Wales in 2008–09, obviously generates more tests but also identifies more cases and therefore places more farms under movement restrictions in the first three years. However, in later years this period of increased testing proves beneficial reducing all four measures below baseline.

Other intervention strategies are aimed at particular transmission routes but their effectiveness is limited as no single route dominates the spread of infection. Decreasing local regional environmental transmission between farms by 50% (dark green), which may represent the potential effect of a large-scale badger cull, has a relatively little effect – we still predict a sustained rise in reactors and farms under restrictions over the entire six years. Similarly, even preventing the movement of all infected animals (dark blue), thereby containing the spread of infection has limited benefit. The introduction of an improved test (light blue), which has higher sensitivity but lower specificity as a follow-up test after a herd breakdown may be compared to the use of gamma-interferon based testing. While reducing the number of herds under restrictions, such a test increases the number of slaughtered animals as more infected animals are detected by this additional test. Finally, although we have identified farms selling many animals as responsible for the majority of onward transmission, targeting such farms with additional standard skin (SICCT) tests is largely ineffective.

Our model highlights the mixture of transmission routes and range of spatial scales involved in the dynamics of bovine TB in Great Britain, and the consequent challenges of eradication. The multifactorial nature of transmission is reflected in our predictions for different interventions; only generic measures such as more national testing, whole herd culling or vaccination that affect all routes of transmission are effective at controlling the spread of bovine TB.

The role of badgers in the maintenance and spread of bovine TB is a matter of considerable scientific, political and public interest^{5,25,26}. Owing to the absence of necessary spatial and population level data on badgers, our model does not explicitly include their role in transmission. The environmental reservoirs play a comparable function, although the contribution of reservoir species and contaminated pasture cannot be separated. The environment is essential in maintaining local infection and may be implicated in up to 80% of all herd breakdowns (Fig. 2b, green hashed area). However, we predict that the environment only contributes weakly to generation of new cases (Fig. 2a), and its effect decays fairly rapidly in the absence of infected cattle (Extended Data Table 1). Therefore, we predict that control of local badger populations and hence control of environmental transmission (Fig. 3, dark green bars) will have a relatively limited effect on all measures of bovine TB incidence.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 September 2012; accepted 27 May 2014.

Published online 2 July 2014.

- Department for Environment Food and Rural Affairs. *Bovine Tuberculosis Evidence Plan* (2013).
- National Audit Office. *Identifying and Tracking Livestock in England* (2003).
- Waddington, K. To stamp out "so terrible a malady": bovine tuberculosis and tuberculin testing in Britain, 1890–1939. *Med. Hist.* **48**, 29–48 (2004).
- Reynolds, D. A review of tuberculosis science and policy in Great Britain. *Vet. Microbiol.* **112**, 119–126 (2006).
- Krebs, J. & The Independent Scientific Review Group. *Bovine Tuberculosis in Cattle and Badgers* (1997).
- Gopal, R., Goodchild, A., Hewinson, G., de la Rua Domenech, R. & Clifton-Hadley, R. Introduction of bovine tuberculosis to north-east England by bought-in cattle. *Vet. Rec.* **159**, 265–271 (2006).
- The Independent Scientific Group on Cattle TB. *Bovine TB: The Scientific Evidence* (2007).
- Department for Environment Food and Rural Affairs. *Bovine TB Eradication Programme for England* <http://www.defra.gov.uk/publications/files/pb13601-bovinetb-eradication-programme-110719.pdf> (2011).
- Pollock, J. M. *et al.* Immune responses in bovine tuberculosis. *Tuberculosis (Edinb.)* **81**, 103–107 (2001).
- Green, L. E. & Cornell, S. J. Investigations of cattle herd breakdowns with bovine tuberculosis in four counties of England and Wales using VETNET data. *Prev. Vet. Med.* **70**, 293–311 (2005).
- Monaghan, M. L., Doherty, M. L., Collins, J. D., Kazda, J. F. & Quinn, P. J. The tuberculin test. *Vet. Microbiol.* **40**, 111–124 (1994).
- Conlan, A. J. K. *et al.* Estimating the hidden burden of bovine tuberculosis in Great Britain. *PLOS Comput. Biol.* **8**, e1002730 (2012).
- Keeling, M. J. *et al.* Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science* **294**, 813–817 (2001).
- Ferguson, N. M., Donnelly, C. A. & Anderson, R. M. The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. *Science* **292**, 1155–1160 (2001).
- Barlow, N. D., Kean, J. M., Hickling, G., Livingstone, P. G. & Robson, A. B. A simulation model for the spread of bovine tuberculosis within New Zealand cattle herds. *Prev. Vet. Med.* **32**, 57–75 (1997).
- Fischer, E. A. J., van Roermund, H. J. W., Hemerik, L., van Asseldonk, M. A. P. M. & de Jong, M. C. M. Evaluation of surveillance strategies for bovine tuberculosis (*Mycobacterium bovis*) using an individual based epidemiological model. *Prev. Vet. Med.* **67**, 283–301 (2005).
- Woolhouse, M. E. J. *et al.* Epidemiological implications of the contact network structure for cattle farms and the 20–80 rule. *Biol. Lett.* **1**, 350–352 (2005).
- Green, D. M., Kiss, I. Z., Mitchell, A. P. & Kao, R. R. Estimates for local and movement-based transmission of bovine tuberculosis in British cattle. *Proc. R. Soc. B* **275**, 1001–1005 (2008).
- Blower, S. M. *et al.* The intrinsic transmission dynamics of tuberculosis epidemics. *Nature Med.* **1**, 815–821 (1995).
- Cox, D. R. *et al.* Simple model for tuberculosis in cattle and badgers. *Proc. Natl Acad. Sci. USA* **102**, 17588–17593 (2005).
- Brooks-Pollock, E. *et al.* Age-dependent patterns of bovine tuberculosis in cattle. *Vet. Res.* **44**, 97 (2013).
- O'Hare, A., Orton, R. J., Bessell, P. R. & Kao, R. R. Estimating epidemiological parameters for bovine tuberculosis in British cattle using a Bayesian partial-likelihood approach. *Proc. R. Soc. B* **281**, 20140248 (2014).
- Sisson, S. A., Fan, Y. & Tanaka, M. M. Sequential Monte Carlo without likelihoods. *Proc. Natl Acad. Sci. USA* **104**, 1760–1765 (2007).
- Toni, T., Welch, D., Strelkowa, N., Ipsen, A. & Stumpf, M. P. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface* **6**, 187–202 (2009).
- Donnelly, C. & Hone, J. Is there an association between levels of bovine tuberculosis in cattle herds and badgers? *Stat. Commun. Infect. Dis.* **2** (2010).
- Godfray, H. C. J. *et al.* A restatement of the natural science evidence base relevant to the control of bovine tuberculosis in Great Britain. *Proc. R. Soc. B* **280**, 20131634 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was funded by BBSRC, the Wellcome Trust and EPSRC. We would like to thank G. Medley, L. Green, O. Courtenay, A. Ramirez-Villaescusa, J. Wood and L. Danon for helpful discussions on bovine TB dynamics. Thanks to A. Conlan and T. J. McKinley for advice on implementing SMC-ABC and to A. Conlan to setting up the Marx Bros cluster. The breakdown and reactor data was supplied by the AHVLA team (particularly A. Mitchell and R. Blackwell), the RADAR team and DEFRA.

Author Contributions M.J.K. and E.B.-P. developed the model structure; E.B.-P. and G.O.R. developed the statistical methodology; all authors contributed to the writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.J.K. (M.J.Keeling@warwick.ac.uk).

Horizontal genome transfer as an asexual path to the formation of new species

Ignacia Fuentes^{1*}, Sandra Stegemann^{1*}, Hieronim Golczyk², Daniel Karcher¹ & Ralph Bock¹

Allopolyploidization, the combination of the genomes from two different species, has been a major source of evolutionary innovation and a driver of speciation and environmental adaptation^{1–4}. In plants, it has also contributed greatly to crop domestication, as the superior properties of many modern crop plants were conferred by ancient allopolyploidization events^{5,6}. It is generally thought that allopolyploidization occurred through hybridization events between species, accompanied or followed by genome duplication^{6,7}. Although many allopolyploids arose from closely related species (congeners), there are also allopolyploid species that were formed from more distantly related progenitor species belonging to different genera or even different tribes⁸. Here we have examined the possibility that allopolyploidization can also occur by asexual mechanisms. We show that upon grafting—a mechanism of plant–plant interaction that is widespread in nature—entire nuclear genomes can be transferred between plant cells. We provide direct evidence for this process resulting in speciation by creating a new allopolyploid plant species from a herbaceous species and a woody species in the nightshade family. The new species is fertile and produces fertile progeny. Our data highlight natural grafting as a potential asexual mechanism of speciation and also provide a method for the generation of novel allopolyploid crop species.

Grafting commonly occurs in nature^{9–13} and the observation of natural grafting may have inspired the application of grafting in agriculture and horticulture hundreds of years BCE¹⁴. Natural stem grafts result from the mechanical pressure of interlocking stems or branches, leading to tissue fusion and establishment of new vascular connections. Natural root grafts^{11,12} may influence population ecology by allowing a community of plants to share water, minerals and metabolites¹⁴. In nature, interspecific grafts have been observed between very distantly related species⁹. Recent work has demonstrated that entire chloroplast genomes can be horizontally transferred across the graft junction, potentially explaining so-called organelle capture events that have occurred in evolution^{15–17}. If nuclear genomes are also exchanged between neighbouring cells in natural grafts, this would provide a straightforward asexual mechanism by which new allopolyploid species could form. Indeed, microscopic evidence supports the idea of nuclear material occasionally moving from cell to cell in intact plant tissues, a phenomenon known as cytomixis^{18–20}.

To investigate the possibility that grafting enables the transfer of nuclear DNA, we generated two transgenic tobacco (*Nicotiana tabacum*) plants carrying different selectable marker genes. Line Nt-kan:yfp carries the kanamycin resistance gene *nptII* (and the gene for the yellow fluorescent reporter YFP) in its nuclear genome, whereas line Nt-hyg contains

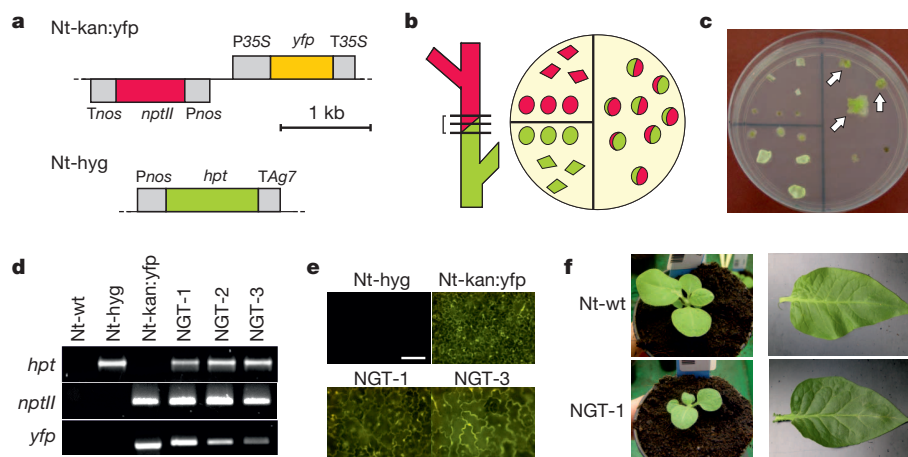


Figure 1 | Grafting and experimental selection for horizontal transfer of nuclear DNA between tobacco plants. **a**, Map of the transgenic loci in the two tobacco lines used for grafting and selection. Genes above the line are transcribed from left to right, genes below the line in the opposite direction. Expression elements (promoters, terminators) are shown in grey. *Pnos*, nopaline synthase promoter; *P35S*, CaMV 35S promoter; *Tnos*, nopaline synthase terminator; *T35S*, CaMV 35S terminator; *TAg7*, *Ag7* terminator from *Agrobacterium tumefaciens*; *nptII*, kanamycin resistance gene; *hpt*, hygromycin resistance gene; *yfp*, yellow fluorescent protein gene. **b**, Schematic drawing of the grafted stem regions, the sectioning of the graft site (horizontal lines) and the selection for gene transfer. The two transgenic lines are coloured according to the antibiotic resistance gene they carry (see **a**). The Petri dish is split in three parts, with the left two quarters containing leaf and stem sections

from the two graft partners and the right half containing sectioned graft sites. **c**, Selection of callus resistant to kanamycin and hygromycin on medium containing both antibiotics. White arrows indicate green resistant tissue. **d**, Presence of all three transgenes in doubly resistant lines (nuclear gene transfer, NGT lines). PCR assays were performed with wild-type tobacco (Nt-wt), the two graft partners (Nt-kan:yfp, Nt-hyg) and three independently generated NGT lines (see Methods). **e**, Analysis of YFP fluorescence by ultraviolet microscopy. Note the differences in cell size between Nt-kan:yfp plants and NGT plants. Scale bar, 100 μ m. **f**, Phenotype of NGT plants. Growth of NGT seedlings is delayed compared to wild-type seedlings (left panel, picture taken 23 days after sowing). Comparison between a wild-type leaf and a leaf from an NGT plant reveals enhanced pigmentation of NGT leaves (right panel).

¹Max-Planck-Institut für Molekulare Pflanzenphysiologie, Am Mühlenberg 1, D-14476 Potsdam-Golm, Germany. ²Department of Molecular Biology, Institute of Biotechnology, John Paul II Catholic University of Lublin, Konstantynów 11, 20-708 Lublin, Poland.

*These authors contributed equally to this work.

the hygromycin resistance gene *hpt* (Fig. 1a). The two lines were grafted onto each other and, after fusion of scion and stock had occurred, the graft site was excised and subjected to double selection for kanamycin and hygromycin resistance (Fig. 1b). Although tissue explants from the two graft partners cannot grow on this medium and bleach out (Fig. 1c), surprisingly, the formation of resistant calli from the graft sites was observed frequently (29 independent events selected from 13 grafts). Doubly resistant tissue readily regenerated into shoots, which were tested for the presence of the two resistance genes and the *yfp* reporter gene (Fig. 1d). PCR assays detected all three transgenes in the selected lines (referred to as NGT lines, for nuclear gene transfer), tentatively suggesting that nuclear gene transfer across the graft junction had occurred (Fig. 1d). As a preliminary investigation to distinguish between transfer of individual genes and transfer of entire nuclear genomes, the macroscopic and microscopic phenotypes of the regenerated plants were examined. Cell size is correlated to nuclear DNA content²¹ and, therefore, polyploid plants often have larger cells than their diploid ancestors. Indeed, NGT plants had larger cells than *N. tabacum* (Fig. 1e), indicating that they contain significantly more DNA. When NGT plants were grown in the greenhouse, they were smaller than wild-type plants and exhibited more intense leaf pigmentation (Fig. 1f). NGT plants were fertile and produced seeds from which the next generation was raised to determine chromosome number and DNA content in individual seedlings. The chromosome set of *N. tabacum* comprises 48 chromosomes. Although 48 chromosomes were present in Nt-kan:yfp and Nt-hyg plants, all T₁ progeny of the NGT lines displayed increased chromosome numbers (Extended Data Fig. 1). The numbers were variable, but a large fraction of the plants contained 96 chromosomes. Hence, the NGT lines are autopolyploid (tetraploid or, considering the allopolyploid origin of *Nicotiana tabacum*, autoallooctoploid). Newly formed autopolyploids may be meiotically unstable and experience the loss of individual chromosomes due to missegregation⁴, potentially explaining why some NGT progeny had fewer than 96 chromosomes. Indeed, when meiotic chromosome segregation was analysed in pollen mother cells, aberrant segregation patterns were observed (Extended Data Fig. 2). Consistent with occasional chromosome loss, the F₁ progeny displayed considerable phenotypic variation (Extended Data Fig. 3). Measurements of nuclear DNA content by flow cytometry confirmed the tetraploid status of NGT plants (and the triploid status of progeny from a cross with the wild type; Extended Data Fig. 1c). Tetraploidy was further confirmed by inheritance tests analysing the segregation of the two antibiotic resistances (Extended Data Table 1).

Grafting and tissue culture procedures can occasionally induce spontaneous polyploidization or formation of chimaeras. Although our inheritance analyses excluded the possibility of NGT plants being chimaeras composed of Nt-kan:yfp cells and Nt-hyg cells, the data do not entirely rule out the remote possibility that the resistance gene was horizontally transferred and the genome duplication occurred independently through spontaneous autopolyploidization in all NGT lines. To definitively prove that entire nuclear genomes are horizontally transferred and to examine whether this transfer provides an asexual mechanism by which new allopolyploid species can form, grafting experiments with two different plant species were performed using the tree tobacco, *Nicotiana glauca* (a woody species), and the cigarette tobacco, *Nicotiana tabacum* (a herbaceous species). *N. glauca* was equipped with the *nptII* gene, whereas *N. tabacum* harboured the *hpt* gene (Fig. 1a). Grafting experiments were performed both *in vitro*¹⁵ and in the greenhouse, mimicking natural stem grafting as occurring in 'kissing trees' (Fig. 2a). From a total of twelve grafted plants, 45 doubly resistant lines were obtained (Fig. 2b) and regenerated into plantlets (Fig. 2c). DNA content measurements revealed that the plants have genome sizes equalling the sum of the *N. glauca* and *N. tabacum* genomes (Fig. 2d). Analysis of molecular markers confirmed the presence of genetic material from both species (Fig. 3), indicating that allopolyploid plants had indeed been obtained through grafting. As plants combining the genomes of *N. tabacum* and *N. glauca* potentially represent a new species, we tentatively named them *Nicotiana tabauca*.

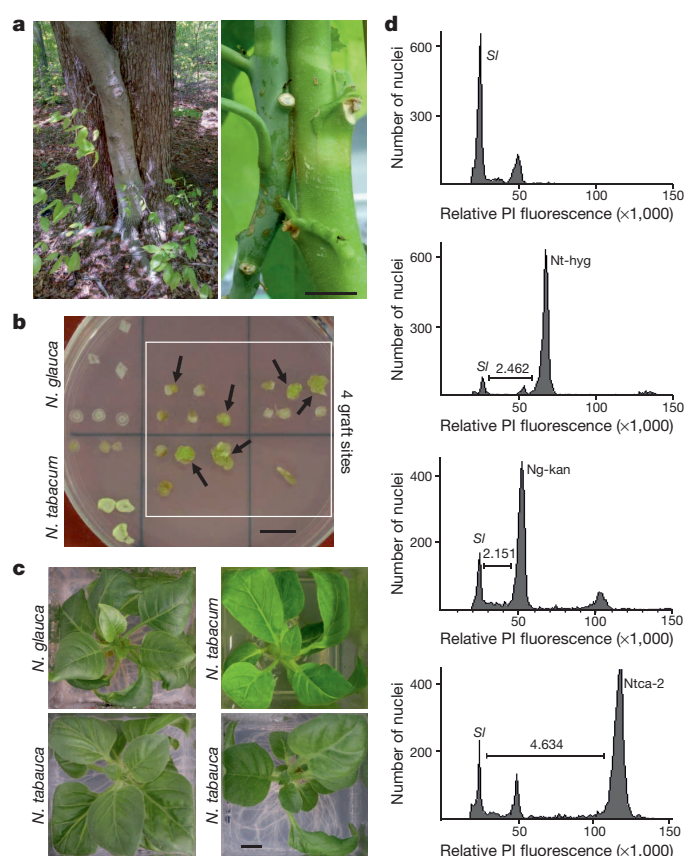


Figure 2 | Grafting and horizontal transfer of nuclear DNA between different species. **a**, A natural stem graft between a beech (front) and a maple (back) in a forest near Monroe, New Jersey (left picture), and an similar stem graft between the tree tobacco, *Nicotiana glauca* (left), and the cigarette tobacco *Nicotiana tabacum* (right) in the greenhouse. **b**, Selection for horizontal transfer of nuclear DNA between plants by exposing graft sites to selection for double resistance to hygromycin and kanamycin. Control explants from the two graft partners (left) and explants from four graft sites (right; framed in white) are shown (compare with Figure 1a–c). **c**, Growth of plantlets regenerated from grafts between *N. tabacum* and *N. glauca* (here termed *Nicotiana tabauca*) on synthetic medium. **d**, Genome size determination in *N. tabauca* by flow cytometry (compare with Extended Data Fig. 1c). *Sl*: *Solanum lycopersicum* (tomato) standard; *Ntca*: *Nicotiana tabauca*. Note that the calculated genome size for *N. tabauca* equals the sum of the genome sizes of *N. tabacum* (Nt-hyg) and *N. glauca* (Ng-kan).

N. tabauca plants combine many traits from their two progenitor species, for example, perennial growth and axillary anthocyanin accumulation from *N. glauca* and the lighter green leaf pigmentation from *N. tabacum*. Other traits are intermediate, such as leaf shape and flower morphology (Fig. 4a, b). Under greenhouse conditions, *N. tabauca* overgrows its two progenitor species (Fig. 4a), possibly indicating that it could outcompete them. The fitness advantage of many natural allopolyploids and the superior yield of allopolyploid crops have been linked to heterosis-like effects^{1,5}.

N. tabauca plants are fertile and produce fertile progeny (Fig. 3b), thus fulfilling all criteria of a new species. Inheritance assays indicate that *N. tabauca* behaves like a true amphipolyploid in that the genomes of the two parental species are sufficiently different to largely prevent pairing of *N. glauca* chromosomes with (homeologous) *N. tabacum* chromosomes. Consequently, nuclear marker genes show disomic inheritance and segregate like in a diploid (Fig. 3b).

Interestingly, the phenotypic variation in the progeny of *N. tabauca* plants was less pronounced than in the autotetraploid NGT lines (Extended Data Fig. 4). None of 24 analysed *N. tabauca* progeny showed a reduction in nuclear DNA content that would be detectable with the (limited)

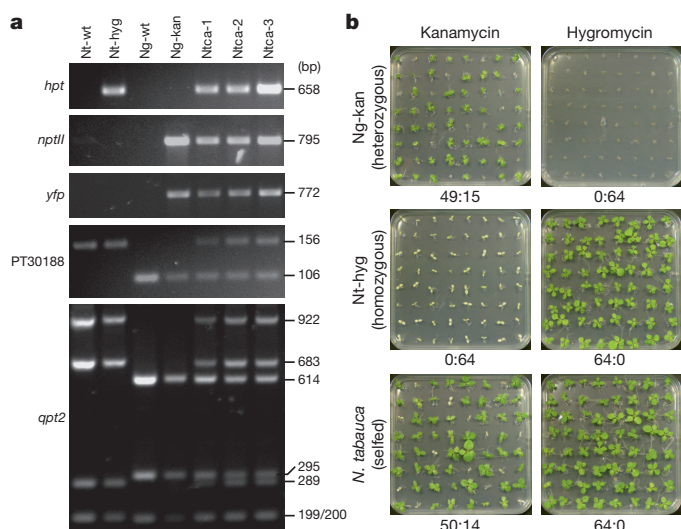


Figure 3 | Presence of nuclear genes from both *Nicotiana tabacum* and *Nicotiana glauca* in *Nicotiana tabauca* and their Mendelian inheritance. **a**, Wild-type *N. tabacum* (Nt-wt), wild-type *N. glauca* (Ng-wt), the two graft partners (Nt-hyg, Ng-kan) and three independently generated *N. tabauca* lines (Ntca-1, 2 and 3) were analysed by PCR using species-specific primers for endogenous genes and transgenes (see Extended Data Table 2 and Methods). PT30188 is a polymorphic microsatellite marker. Amplification products of the *qpt2* gene were digested with a restriction enzyme (for details and fragments sizes, see Methods). **b**, Segregation analysis of *N. tabauca* in comparison to its two parental species, *N. glauca* (Ng-kan) and *N. tabacum* (Nt-hyg). Segregation of the two antibiotic resistances is shown for the two graft partners (a heterozygous Ng-kan and a homozygous Nt-hyg) and a selfed *N. tabauca* line. Segregation ratios are given as numbers of antibiotic-resistant to antibiotic-sensitive seedlings. Note that the transgenes show disomic inheritance and segregate like in a diploid. Consequently, as the *N. tabacum* parent was homozygous for *hpt*, no hygromycin-sensitive progeny were obtained from selfed *N. tabauca*.

sensitivity of flow cytometry. However, karyotype analyses in root meristems revealed, in addition to plants with the full allopolyploid set of chromosomes (Fig. 4c), also cases of chromosome loss (Extended Data Fig. 5), suggesting that there is some somatic genome instability, at least in rapidly dividing root tip cells. Thus, full evolutionary stabilization of the new allopolyploid species will require further genomic adaptations, which occur rapidly after polyploidization events^{22,23}. Cases of chromosome loss and genome instability have been documented in both recently evolved natural allopolyploids and synthetic allopolyploids^{24,25}, indicating that genomic imbalances are a common consequence of allopolyploidization.

The movement of entire nuclear genomes from cell to cell and across graft junctions raises questions about the underlying transfer mechanism. Two mechanisms appear plausible: (1) fusion of neighbouring cells at the graft site, or (2) migration of nuclei from cell to cell through plasmodesmata in a cytomixis-like process, as observed in microscopic studies^{18–20}. To obtain insights into the transfer mechanism, we investigated the genomes of plastids and mitochondria. These organelles occur at high numbers per cell and, if cell fusion was involved, one might expect *N. tabauca* plants to harbour a mixed population of organelles from *N. glauca* and *N. tabacum*, a genetic heterogeneity known as heteroplasmy. Analysis of several independently generated *N. tabauca* plants revealed that some contained the *N. glauca* plastid genome, whereas others contained the *N. tabacum* plastid genome, but none was heteroplasmic (Extended Data Fig. 6), arguing against cell fusion as the mechanism underlying genome transfer. By contrast, some heteroplasmy was detected for the mitochondrial genome (Extended Data Figs 7 and 8). However, as mitochondria are often tightly associated with the nuclear membrane and sometimes even found within the nucleus²⁶, this finding does not necessarily provide evidence against cytomixis. Alternatively, segregation of plastids could occur faster than segregation of mitochondria.

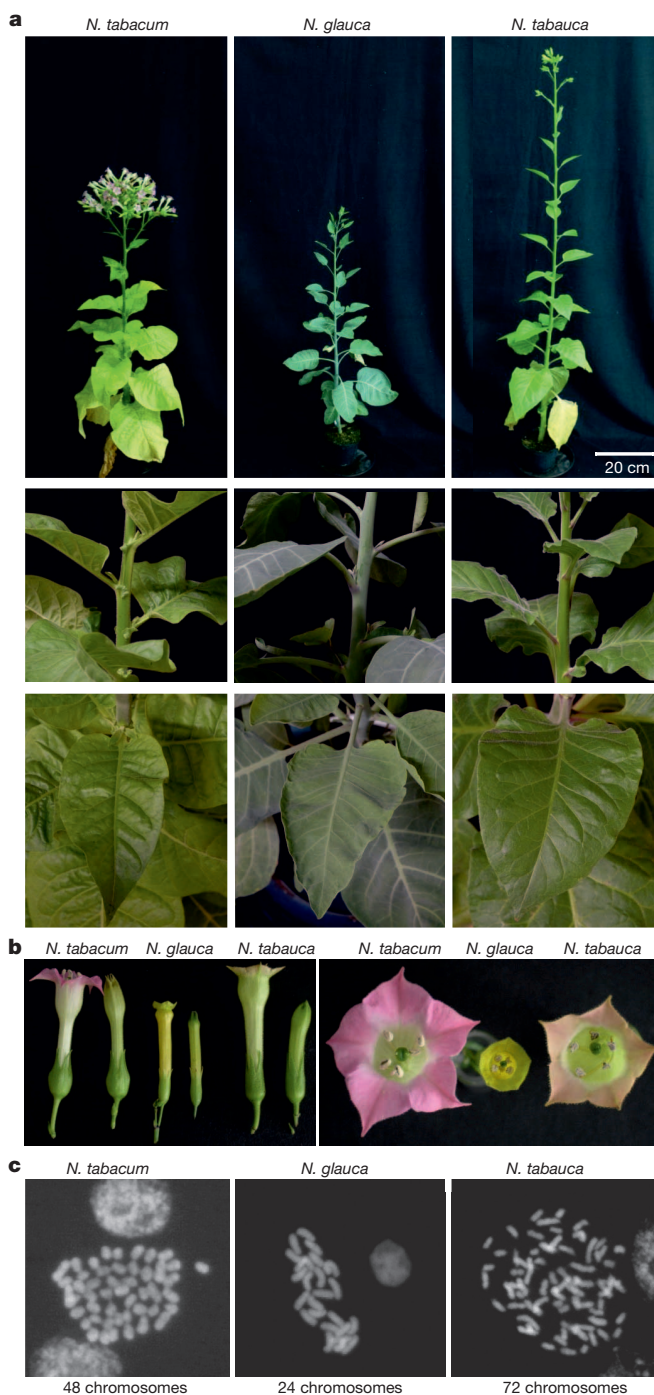


Figure 4 | Phenotype and karyotype of *Nicotiana tabauca* plants. **a**, Vegetative phenotypes of *N. tabacum*, *N. glauca* and *N. tabauca*. Plants in the upper panel were photographed 44 days after transfer to soil. Note the difference in leaf colour (blue-green in *N. glauca* and lighter green in *N. tabacum* and *N. tabauca*). The middle panel shows the difference in anthocyanin accumulation. *N. glauca* has accumulation visible in the whole leaf; *N. tabauca* has accumulation restricted to the petiole base; *N. tabacum* has no anthocyanin. The lower panel depicts the difference in leaf morphology. **b**, Comparison of flower morphology. **c**, Karyotype analysis by DAPI (4',6-diamidino-2-phenylindole) staining reveals the allopolyploid status of *N. tabauca*.

The initial interspecies genome transfer in a natural graft represents a somatic event. However, it can become heritable by lateral shoot formation from the graft site, which is facilitated by the presence of a meristematic ring surrounding the stem, the cambium¹². Lateral shoot formation

is commonly induced by wounding, for example, upon herbivory or grafting^{10,14}. This would allow the allopolyploid cell to enter the germline by becoming part of a newly formed apical meristem.

As most modern crop plants arose through allopolyploidization, the possibility to generate new allopolyploid species by grafting-mediated genome transfer also provides a new tool for crop improvement. Although genome combination can also be achieved synthetically using protoplast fusion techniques^{27,28}, the procedures involved are technically demanding and available only for a limited number of species. In cases where hybridization is possible, the genomes of two species can also be combined sexually, but the resulting hybrids are often sterile unless their chromosome sets are doubled. This is also the case for most interspecific hybrids between *Nicotiana* species, including *N. tabacum* and *N. glauca*²⁹. Finally, genome transfer by grafting creates new combinations of plastid and mitochondrial genomes (Extended Data Figs 6–8). This is important because the organellar genotypes influence important agronomic traits, including growth and stress tolerance³⁰.

We have demonstrated that grafting results in the transfer of entire nuclear genomes between species. The significance of this is twofold. First, natural grafting provides a potential asexual mechanism of speciation that is much less restricted by incompatibilities than sexual hybridization^{9,14} (Fig. 2a). Thus, grafting should be considered as an alternative mechanism of polyploidization, especially in allopolyploids that formed from distantly related species. Second, interspecies grafting provides a method to produce new allopolyploid crop species. As polyploidy confers the superior properties of modern crops over their diploid progenitor species^{6,7}, this has significant potential in breeding and agricultural biotechnology.

METHODS SUMMARY

Nicotiana plants were grown under greenhouse conditions or under aseptic conditions on synthetic medium. Transgenic plants were produced by *Agrobacterium*-mediated transformation. Interspecies grafting experiments were performed reciprocally by transplanting *N. glauca* scions onto *N. tabacum* stocks and vice versa. Selection for gene transfer was done on regeneration medium containing kanamycin and hygromycin. Regenerated doubly resistant shoots were rooted on synthetic medium, transferred to soil and grown to maturity in the greenhouse. For molecular analyses, microscopy, karyotyping and flow cytometry, see the Methods section.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 November 2013; accepted 31 March 2014.

Published online 8 June 2014.

- Comai, L. The advantages and disadvantages of being polyploid. *Nature Rev. Genet.* **6**, 836–846 (2005).
- Sobel, J. M., Chen, G. F., Watt, L. R. & Schemske, D. W. The biology of speciation. *Evolution* **64**, 295–315 (2010).
- Madlung, A. Polyploidy and its effect on evolutionary success: old questions revisited with new tools. *Heredity* **110**, 99–104 (2013).
- Leitch, A. R. & Leitch, I. J. Genomic plasticity and the diversity of polyploid plants. *Science* **320**, 481–483 (2008).
- Paterson, A. H. Polyploidy, evolutionary opportunity, and crop adaptation. *Genetica* **123**, 191–196 (2005).
- Hegarty, M. J. & Hiscock, S. J. Genomic clues to the evolutionary success of polyploid plants. *Curr. Biol.* **18**, R435–R444 (2008).

- Soltis, P. S. & Soltis, D. E. The role of hybridization in plant speciation. *Annu. Rev. Plant Biol.* **60**, 561–588 (2009).
- Joly, S., Heenan, P. B. & Lockhart, P. J. A Pleistocene inter-tribal allopolyploidization event precedes the species radiation of *Pachycladon* (Brassicaceae) in New Zealand. *Mol. Phylogenet. Evol.* **51**, 365–372 (2009).
- Seidel, C. F. Ueber Verwachsungen von Stämmen und Zweigen von Holzgewächsen und ihren Einfluss auf das Dickenwachstum der betreffenden Theile. *Naturwiss. Ges. Isis Dresden Sitzber.* 161–168 (1879).
- Küster, E. Über Stammverwachsungen. *Jahrb. Wiss. Bot.* **33**, 487–512 (1899).
- Beddie, A. D. Natural root grafts in New Zealand trees. *Transact. Proc. R. Soc. New Zeal.* **71**, 199–203 (1942).
- Larson, P. R. The vascular cambium: development and structure. *Springer Series in Wood Science* 1–725 (Springer-Verlag, 1994).
- Bock, R. The give-and-take of DNA: horizontal gene transfer in plants. *Trends Plant Sci.* **15**, 11–22 (2010).
- Mudge, K., Janick, J., Scofield, S. & Goldschmidt, E. E. A history of grafting. *Hortic. Rev. (Am. Soc. Hortic. Sci.)* **35**, 437–493 (2009).
- Stegemann, S. & Bock, R. Exchange of genetic material between cells in plant tissue grafts. *Science* **324**, 649–651 (2009).
- Stegemann, S., Keuthe, M., Greiner, S. & Bock, R. Horizontal transfer of chloroplast genomes between plant species. *Proc. Natl Acad. Sci. USA* **109**, 2434–2438 (2012).
- Thyssen, G., Svab, Z. & Maliga, P. Cell-to-cell movement of plastids in plants. *Proc. Natl Acad. Sci. USA* **109**, 2439–2443 (2012).
- Zhang, W. C., Yan, W. M. & Lou, C. H. Intercellular movement of protoplasm *in vivo* in developing endosperm of wheat caryopses. *Protoplasma* **153**, 193–203 (1990).
- Zhang, W.-C. Progress in research on intercellular movement of protoplasm in higher plants. *Acta Bot. Sin.* **44**, 1068–1074 (2002).
- Mursalimov, S. R., Sidorchuk, Y. V. & Deineko, E. V. New insights into cytomixis: specific cellular features and prevalence in higher plants. *Planta* **238**, 415–423 (2013).
- Melaraño, J. E., Mehrotra, B. & Coleman, A. W. Relationship between endopolyploidy and cell size in epidermal tissue of *Arabidopsis*. *Plant Cell* **5**, 1661–1668 (1993).
- Madlung, A. et al. Genomic changes in synthetic *Arabidopsis* polyploids. *Plant J.* **41**, 221–230 (2005).
- Pontes, O. et al. Chromosomal locus rearrangements are a rapid response to formation of the allotetraploid *Arabidopsis suecica* genome. *Proc. Natl Acad. Sci. USA* **101**, 18240–18245 (2004).
- Chester, M. et al. Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). *Proc. Natl Acad. Sci. USA* **109**, 1176–1181 (2012).
- Xiong, Z., Gaeta, R. T. & Pires, J. C. Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. *Proc. Natl Acad. Sci. USA* **108**, 7908–7913 (2011).
- Yu, H.-S. & Russell, S. D. Occurrence of mitochondria in the nuclei of tobacco sperm cells. *Plant Cell* **6**, 1477–1484 (1994).
- Shepard, J. F., Bidney, D., Barsby, T. & Kemble, R. Genetic transfer in plants through interspecific protoplast fusion. *Science* **219**, 683–688 (1983).
- Evans, D. A., Wetter, L. R. & Gamborg, O. L. Somatic hybrid plants of *Nicotiana glauca* and *Nicotiana tabacum* obtained by protoplast fusion. *Physiol. Plant.* **48**, 225–230 (1980).
- Trojak-Goluch, A. & Berbec, A. Cytological investigations of the interspecific hybrids of *Nicotiana tabacum* L. × *N. glauca* Grah. *J. Appl. Genet.* **44**, 45–54 (2003).
- Greiner, S. & Bock, R. Tuning a ménage à trois: co-evolution and co-adaptation of nuclear and organellar genomes in plants. *Bioessays* **35**, 354–365 (2013).

Acknowledgements We thank the MPI-MP Green Team for help with plant transformation, K. Köhl for providing the *hpt* vector and S. Ruf (MPI-MP) for discussions. We are grateful to J. Fuchs (IPK Gatersleben) for advice on flow cytometry measurements. This research was financed by the Max Planck Society.

Author Contributions I.F., S.S. and H.G. performed the experiments. All authors participated in data evaluation and experimental design. R.B. conceived the study, and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.B. (rbock@mpimp-golm.mpg.de).

Cntnap4 differentially contributes to GABAergic and dopaminergic synaptic transmission

T. Karayannis^{1*}, E. Au^{1*}, J. C. Patel², I. Kruglikov¹, S. Markx³, R. Delorme^{4,5,6}, D. Héron⁷, D. Salomon⁸, J. Glessner⁹, S. Restituto¹, A. Gordon⁸, L. Rodriguez-Murillo³, N. C. Roy^{1,10}, J. A. Gogos¹¹, B. Rudy¹, M. E. Rice², M. Karayiorgou³, H. Hakonarson⁴, B. Keren¹², G. Huguet^{4,5,13}, T. Bourgeron^{4,5,13,14}, C. Hoeffler¹, R. W. Tsien¹, E. Peles³ & G. Fishell¹

Although considerable evidence suggests that the chemical synapse is a lynchpin underlying affective disorders, how molecular insults differentially affect specific synaptic connections remains poorly understood. For instance, Neurexin 1a and 2 (*NRXN1* and *NRXN2*) and *CNTNAP2* (also known as *CASPR2*), all members of the neurexin superfamily of transmembrane molecules, have been implicated in neuropsychiatric disorders. However, their loss leads to deficits that have been best characterized with regard to their effect on excitatory cells^{1,2}. Notably, other disease-associated genes such as *BDNF* and *ERBB4* implicate specific interneuron synapses in psychiatric disorders^{3,4}. Consistent with this, cortical interneuron dysfunction has been linked to epilepsy, schizophrenia and autism^{5,6}. Using a microarray screen that focused upon synapse-associated molecules, we identified *Cntnap4* (contactin associated protein-like 4, also known as *Caspr4*) as highly enriched in developing murine interneurons. In this study we show that *Cntnap4* is localized presynaptically and its loss leads to a reduction in the output of cortical parvalbumin (PV)-positive GABAergic (γ -aminobutyric acid producing) basket cells. Paradoxically, the loss of *Cntnap4* augments midbrain dopaminergic release in the nucleus accumbens. In *Cntnap4* mutant mice, synaptic defects in these disease-relevant neuronal populations are mirrored by sensory-motor gating and grooming endophenotypes; these symptoms could be pharmacologically reversed, providing promise for therapeutic intervention in psychiatric disorders.

Having detected *Cntnap4* in developing cortical interneurons, we examined its expression at later stages of development. *In situ* hybridization revealed that it is widely but sparsely distributed in a location pattern similar to cortical interneurons. Double *in situ* hybridization demonstrated that almost all *Cntnap4*-positive cells within the somatosensory cortex are GAD67 (also known as GAD1)-positive (Fig. 1a and Extended Data Fig. 1b). Using a *Cntnap4*-eGFP knock-in allele of eGFP driven from the *Cntnap4* locus (Extended Data Fig. 1a), we assessed its expression in cortical interneuron subtypes. Analysis at postnatal day 21 (P21) revealed that many eGFP⁺ neurons also expressed PV (47 \pm 4%). Conversely, 64 \pm 3% of all PV⁺ interneurons were eGFP⁺ (Fig. 1b and Extended Data Fig. 1d, e). The remaining 53% of *Cntnap4*-eGFP⁺ neurons were immunopositive for other interneuron markers, reelin (RELN), VIP, NPY and calretinin, but not somatostatin (Extended Data Fig. 1c). As the mice matured, there was a steady rise in the percentage of *Cntnap4*-eGFP⁺ PV cells. By P60 almost all PV cells expressed *Cntnap4* (94 \pm 3%; PV⁺ cells as a proportion of total *Cntnap4*-eGFP⁺ cells was 76 \pm 2%) (Fig. 1b), suggesting a possible involvement of *Cntnap4* in their maturation.

In addition, *Cntnap4* expression was also enriched in the substantia nigra pars compacta (SNc) and ventral tegmental area (VTA) midbrain dopaminergic projection populations. Approximately 90% of the tyrosine hydroxylase (TH)-positive dopaminergic neurons also expressed *Cntnap4* (Fig. 1c).

Previously, *Cntnap4* was shown to associate with presynaptic proteins such as NB-2 (also known as contactin 5), as well as Mint1 (also known as Apb1) and CASK, which are important for inhibitory synapses^{7–10}. We examined pre- and postsynaptic fractions prepared from wild type and knockout *Cntnap4* animals. *Cntnap4* was found to be enriched in synapses, exclusively within the presynaptic compartment (Fig. 1d). Additional verification of these results was obtained by using a fusion of extracellular *Cntnap4* and Fc domains on hippocampal neuronal cultures, demonstrating the majority of *Cntnap4* protein is found on the cell body and proximal dendrites in a punctate pattern, co-localized with gephyrin (Fig. 1e and Extended Data Fig. 1f). Hence, *Cntnap4* is highly expressed in cortical PV cells and midbrain dopaminergic neurons and is localized presynaptically.

To assess *Cntnap4* function, we generated two lines of *Cntnap4* null mice of mixed inbred background (line no. 149 and line no. 13) by replacing the first coding exon of *Cntnap4* with eGFP and a neomycin cassette (Extended Data Fig. 1a). Focusing primarily on the 149 line, we examined the functional consequences of *Cntnap4* loss on the dopaminergic and GABAergic populations. We used fast-scan cyclic voltammetry to monitor axonal dopamine spillover in the caudate putamen (CPU) and nucleus accumbens (NAc) of heterozygous, knockout and wild-type mice¹¹. Dopamine release was evoked by single or brief pulse trains (20 pulses at 10 Hz or 5 pulses at varying frequencies). Both heterozygous and knockout mice compared to wild-type animals, showed an increase in peak extracellular dopamine concentration ([DA]_o) that was more pronounced in the NAc than CPU, especially with multiple pulse stimulation (Fig. 2a). Moreover, evoked [DA]_o was enhanced evenly across frequencies varying from 5 to 100 Hz (Extended Data Fig. 2a) indicating both tonic and phasic firing are affected. Analysis of the maximum dopamine uptake rate, V_{max}^{12} , revealed that the rates were somewhat higher in heterozygous and knockout animals compared to wild type (CPU: 7.48 \pm 0.30 μ M s⁻¹ mutants versus 7.25 \pm 0.43 μ M s⁻¹ in wild type; NAc: 5.39 \pm 0.23 μ M s⁻¹ mutants versus 4.95 \pm 0.23 μ M s⁻¹ in wild type). However, the excitability of dopaminergic neurons between knockout and wild-type cells appeared unaffected (data not shown). Hence, increased evoked [DA]_o in *Cntnap4* mutants does not appear to be a consequence of impaired dopamine transporter activity or intrinsic

¹Department of Neuroscience and Physiology, NYU Neuroscience Institute, New York, New York 10016, USA. ²Department of Neurosurgery, Neuroscience and Physiology, New York University Langone Medical Center, New York, New York 10016, USA. ³Department of Psychiatry, College of Physicians and Surgeons, Columbia University, 1051 Riverside Drive, New York, New York 10032, USA. ⁴Institut Pasteur, Human Genetics and Cognitive Functions Unit, 75724 Paris, France. ⁵CNRS URA 2182 Genes, Synapses and Cognition, Institut Pasteur, 75724 Paris, France. ⁶Assistance Publique-Hôpitaux de Paris, Robert Debré Hospital, Department of Child and Adolescent Psychiatry, 75019 Paris, France. ⁷Unité Fonctionnelle de Génétique Médicale AP-HP, Département de Génétique et Cytogénétique, Centre de Référence, Déficiences intellectuelles de causes rares, CRICM, UMR-S975, Groupe Hospitalier Pitié-Salpêtrière, F-75013, Paris, France. ⁸Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel. ⁹Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. ¹⁰Section on Synaptic Transmission, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, Maryland 20892, USA. ¹¹Departments of Physiology and Cellular Biophysics and Neuroscience, Columbia University Medical Center, New York, New York 10032, USA. ¹²Unité Fonctionnelle de Génétique Chromosomique AP-HP, Département de Génétique et Cytogénétique, CRICM, UMR-S975, Groupe Hospitalier, Pitié-Salpêtrière, 75013 Paris, France. ¹³University Paris Diderot, Sorbonne Paris Cité, Human Genetics and Cognitive Functions, 75005 Paris, France. ¹⁴FondaMental Foundation, 94000 Créteil, France.

*These authors contributed equally to this work.

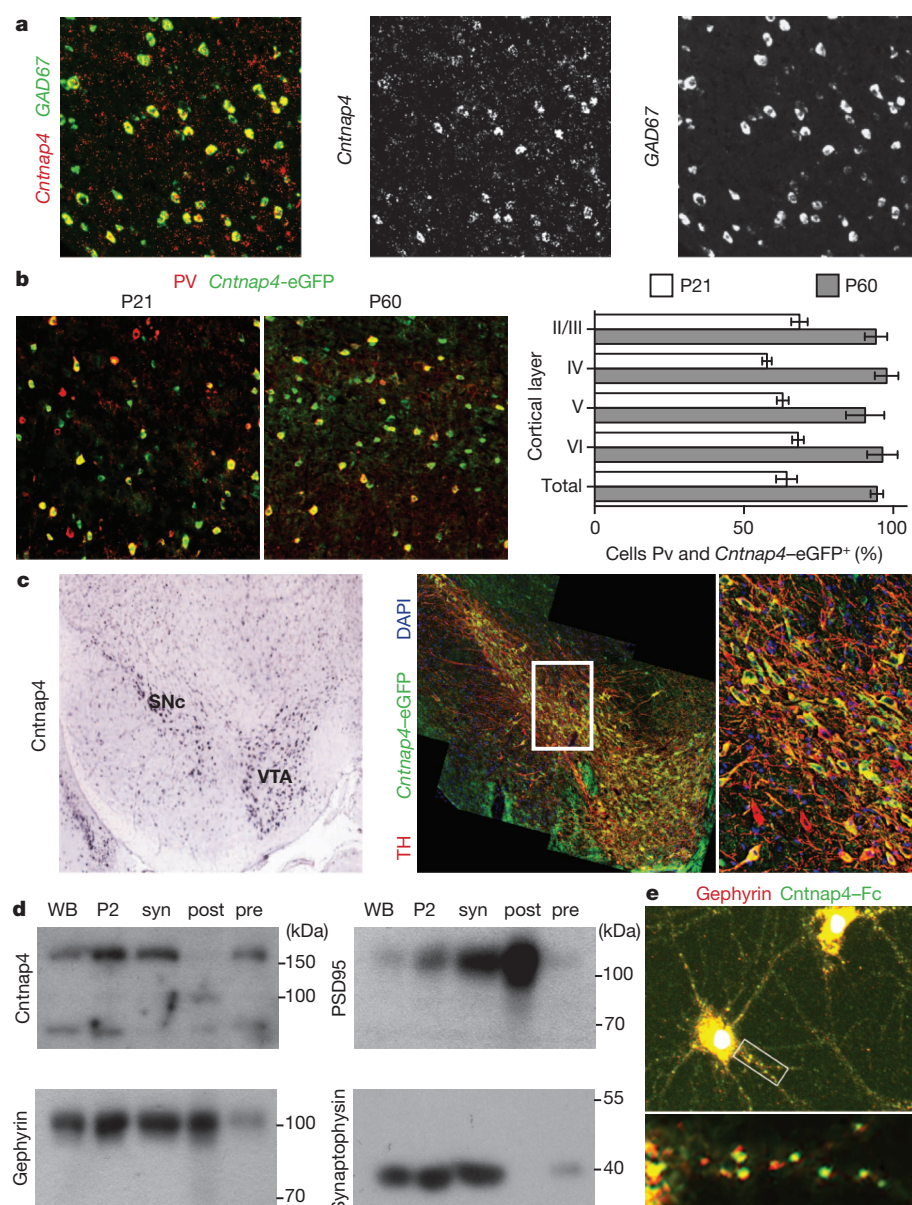


Figure 1 | Cellular and subcellular localization of *Cntnap4*. **a**, Double fluorescent *in situ* hybridization on P21 wild-type somatosensory cortex with co-localization of *Cntnap4* (green) and *GAD67* (red). **b**, Overlap between parvalbumin (PV) (red) and *Cntnap4* (eGFP knock-in) at P21 and P60. **c**, *In situ* hybridization of *Cntnap4* in VTA and SNc of adult WT (©2014 Allen Institute for Brain Science. Allen Mouse Brain Atlas [Internet]. Available from: <http://mouse.brain-map.org/>) and *Cntnap4*-eGFP (green) co-localization with tyrosine hydroxylase (red) at P21. Close-up of the area within the white box is shown on the right. **d**, Western blots on whole brain (WB), total membrane (P2), synaptosomes (syn), presynaptic (pre) and postsynaptic (post) compartments. *Cntnap4* present at presynaptic site (pre). Internal controls for pre- (synaptophysin) and post-synaptic (gephyrin and PSD95) fractions (3 biological replicates, $n = 3$ brains each replicate). **e**, *Cntnap4*-Fc fusion protein (green) binds to soma and proximal dendrites, apposed to gephyrin puncta (red) in dissociated cortical neurons (3 biological replicates, $n = 10$ coverslip cultures analysed).

electrophysiological properties. We conclude that *Cntnap4* normally acts to attenuate dopamine release through a presynaptic mechanism.

Given the dopamine findings, we expected that GABAergic signalling would also be elevated in mutant mice. However, in *Cntnap4* knockout animals compared to controls, spontaneous inhibitory postsynaptic currents (sIPSCs) in pyramidal cells were fewer, smaller and slower (Extended Data Fig. 2b). Intriguingly, paired-cell recordings between PV and excitatory cells in three-week-old knockout mice had synaptic responses reminiscent of those in immature fast spiking (FS) cells¹³ (Fig. 2b). IPSC amplitude was reduced and kinetics were prolonged, with longer rise-times and decay tau values (Fig. 2b, c and Extended Data Fig. 2c). In addition, the average latency of the IPSCs was marginally increased and between-trials variability (jitter) was larger (Fig. 2b, c). Some of these defects persisted into adulthood and hence were not due to developmental delay (Extended Data Fig. 3a). By comparing *Cntnap4*-positive and *Cntnap4*-negative PV interneurons at P21 (Fig. 2c) we observed that negative cells resembled mutant neurons. This indicates not only that the defects we saw are cell-autonomous (*Cntnap4*-negative PV-positive interneurons, grey symbols versus *Cntnap4* knockout cells, red symbols, Fig. 2c), but also that *Cntnap4* appears necessary for the full maturation of the output of PV-positive interneurons.

We then sought to address the mechanism by which the loss of *Cntnap4* affects the function of PV interneurons. A change in the probability of release seems unlikely as the paired pulse ratios in heterozygous and knockout PV interneurons were unchanged (Extended Data Fig. 2d). Moreover, no obvious differences in intrinsic firing properties, morphology or synaptogenesis between wild type, heterozygous and knockout animals (Extended Data Fig. 4a–c) were observed. Another possible mechanism is a failure to transition from N-type to P/Q-type presynaptic calcium channels^{13,14}. However, application of the specific N-type calcium channel blocker ω -conotoxin GVIA in knockout paired recordings ($n = 3$) produced no change in postsynaptic responses (Extended Data Fig. 2e). Yet another possibility is that *Cntnap4* affects the localization of postsynaptic GABA_A receptors. Protein samples from P60 *Cntnap4* wild-type and knockout cortex were sub-fractionated and tested by western blot for levels of GABA_A- $\gamma 2$ (probes for synaptically localized GABA_A receptors) as well as GABA_A- $\alpha 1$ (enriched in FS-pyramidal synapses) and, if anything, we observed a slightly elevated level for both receptor subunits in *Cntnap4* knockout samples (Extended Data Fig. 4d). As *Cntnap* proteins help localize glycosyl phosphatidylinositol (GPI)-anchored cell adhesion molecules¹⁵ and neuroligins¹⁶ have a structural adhesion role at the synapse, we examined symmetric perisomatic

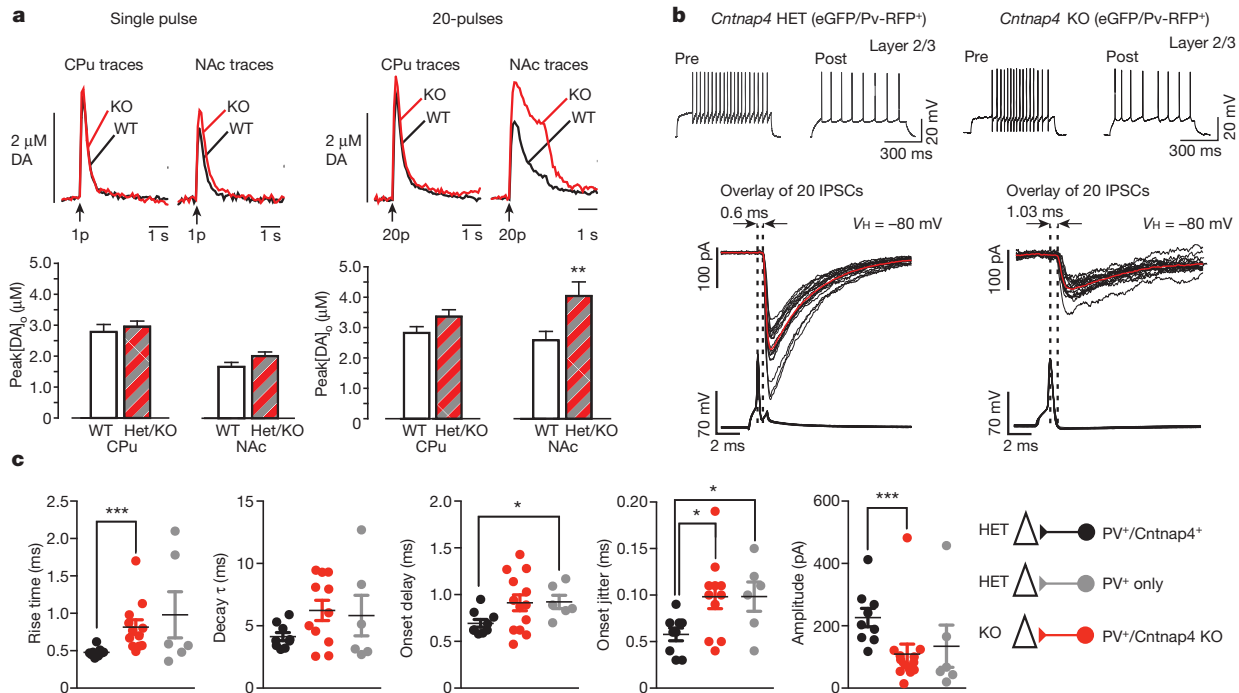
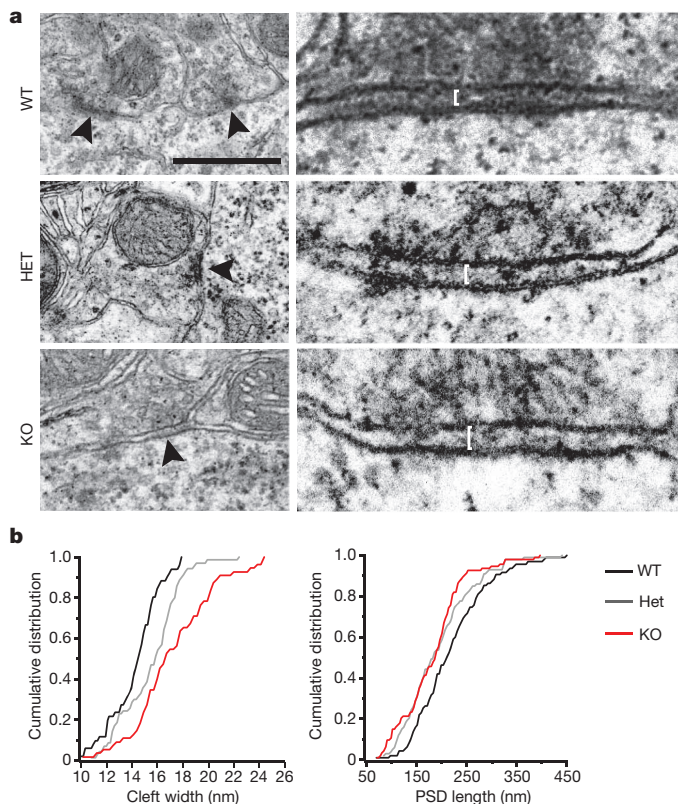


Figure 2 | *Cntnap4* mutant mice show increased dopamine but decreased GABA signalling. **a**, Evoked dopamine measurements by voltammetry *in vitro* in wild type and *Cntnap4* mutants. Increased extracellular dopamine release in mutants versus wild type. Twenty pulses lead to a significant increase in NAc. Representative traces of extracellular dopamine concentration levels [DA]_o over time of wild type (black) versus knockout mice (red). Distribution of peak [DA]_o values for all the data points included in the analysis ($n = 3$ for each genotype). **b**, Paired recordings between PV-positive and excitatory cells in somatosensory cortex *in vitro*. Examples of firing in pre- and postsynaptic cells

upon step depolarization. A single action potential evokes fast, reliable postsynaptic responses in heterozygotes. Responses were smaller, slower and more unreliable in knockouts (individual traces are shown in grey and the average shown in red). **c**, Overall data of synaptic values for responses recorded in three groups of paired recordings. Black denotes presynaptic *Cntnap4*-positive PV cells in heterozygous mice; grey denotes PV cells yet to express *Cntnap4* in heterozygous mice and red denotes PV cells in knockout mice ($n = 7$ brains for each HET and KO; $n = 9$ for black, $n = 13$ for red, $n = 6$ for grey paired recordings; Mann–Whitney *U*-test).



synapses onto pyramidal cells by electron microscopy (EM) in adult knockout, heterozygous and wild-type animals (Fig. 3a). We observed that the cleft width was significantly larger in the heterozygotes versus wild types, and was further increased in the knockout mice (wild type, 14.37 ± 0.28 nm; heterozygote, 15.61 ± 0.28 nm; knockout, 17.37 ± 0.43 nm) (Fig. 3a, b), whereas the postsynaptic density (PSD) was shortest in the knockout and intermediate in heterozygotes compared to wild-type mice (Fig. 3a, b) (wild-type, 221.71 ± 7.34 nm, heterozygote, 195.45 ± 6.9 nm, knockout, 185.34 ± 6.67 nm). By contrast, excitatory (compared to inhibitory) synapses exhibited a significant but milder widening of the synaptic cleft in knockout versus wild-type mice (6.4% versus a 20.8% increase; Extended Data Fig. 5) and no difference in excitatory PSD length. Our results show that *Cntnap4* joins the short list of molecules that contribute to the structural maturation of inhibitory interneuron synapses^{3,4,17,18} and acts in a gene dose-dependent manner¹⁹.

A number of synaptic molecules, including *Cntnap2*, have been extensively implicated in neuropsychiatric disorders. Given the synaptic localization of *Cntnap4* and its differential effects on synaptic transmission in dopaminergic versus GABAergic neurons, it is a promising disease

Figure 3 | Loss of *Cntnap4* results in ultrastructural deficits in perisomatic inhibitory synapses. **a**, Electron micrographs of symmetric perisomatic synaptic contacts (black arrowheads) in wild type (WT), heterozygous (Het) and knockout (KO) littermate mice. Higher magnification shown on right (white bars indicate cleft width). Scale bar, 500 nm. **b**, Cumulative distributions of cleft width and PSD length. (PSD width of WT versus KO $P < 0.0001$; WT versus HET $P = 0.002$; HET versus KO $P = 0.005$; PSD length WT versus KO $P = 0.007$; WT versus HET $P = 0.049$; HET versus KO $P = 0.399$; $n = 3$ brains for each genotype; WT synapse numbers: $n = 51$ for width and $n = 95$ for length, HET synapse numbers: $n = 70$ for width and $n = 99$ for length, KO synapse numbers: $n = 55$ for width and $n = 94$ for length; Kolmogorov–Smirnov test.)

target. Indeed, in addition to a handful of previously reported cases (see Extended Data Fig. 6a and Supplementary Information) our human genetics analysis identified 8 individuals with psychiatric illness and *CNTNAP4* gene disruption. Two harboured coding deletions, with one missing the whole gene, the other lacking the last 3 exons. The remaining six individuals had non-coding deletions, all contained within the 5' region of intron II (Extended Data Figs 6–8 and Supplementary Information). Given these findings and the behavioural abnormalities described in *Cntnap2* mutant animals¹, we tested whether *Cntnap4* heterozygous or knockout mice exhibited abnormal behaviours consistent with neuropsychiatric disorders: grooming, pre-pulse inhibition (PPI), marble burying and behaviour in an open field arena (OFA) and an elevated plus maze (EPM).

Repetitive, perseverative movements comprise a common behavioural abnormality in individuals with autism spectrum disorders (ASDs) and manifest in mice as over-grooming^{20,21}. Both knockout and heterozygous *Cntnap4* mice displayed a severe and highly penetrant over-grooming behaviour, resulting in whisker, face and sometimes body hair loss but rarely lesions (Fig. 4a). This was observed equally in male and female mice, and was apparent before weaning (Fig. 4b). It was even evident in wild-type offspring raised by mutant parents. By cross-fostering, we established unequivocally that allo-grooming in affected litters was always associated with the presence of a mutant *Cntnap4* allele (heterozygous or knockout) in one or both of the parents (Extended Data Fig. 9g). Strain differences also affected this behaviour, as line no. 13 did not display robust over-grooming when crossed onto an outbred background, but regained it when re-crossed into an inbred strain (Supplementary Information). Moreover, the observed over-grooming behaviour in mutant mice is unlikely to stem from overt changes in anxiety levels or deficits in locomotion, as OFA, EPM and marble burying tests did not reveal significant differences between the three genotypes (Extended Data Fig. 9a–f).

In addition to perseverative behaviours, neuropsychiatric patients often show impaired ability to process sensory information²². We performed PPI of the auditory startle reflex^{22,23} and found that heterozygous and knockout mice exhibited both elevated startle responses and abnormal

PPI indexes (Fig. 4c, d). On the basis of our cellular findings, we reasoned that the behavioural phenotypes might be reversed by augmenting inhibitory output and by dampening dopaminergic signalling.

There is extensive literature linking increased activity in the dopaminergic system and overt repetitive behaviours in both mice and humans^{24,25}. We therefore administered the D2 receptor antagonist haloperidol via slow-release pellets (0.2 mg per kg or 0.6 mg per kg) implanted subcutaneously²⁶. Haloperidol-treated parents showed a significant reduction in their own grooming score and did not over-groom their pups (Fig. 4f). In contrast, the pairs treated with vehicle did not show any overall hair and/or whisker recovery and continued to over-groom themselves and their offspring (Fig. 4f). After the 90-day treatment period when haloperidol levels were depleted, over-grooming progressively re-emerged (Fig. 4f, g). The effect of haloperidol on grooming was not the result of hypoactivity (Extended Data Fig. 10a). Therefore, increased dopaminergic signalling in mutant *Cntnap4* mice *in vivo* leads to over-grooming, a behaviour that can be rescued by chronic pharmacological treatment.

In order to assess if the defects in GABAergic transmission can account for the heightened startle or PPI defect in *Cntnap4* mutant animals, indiplon was administered. Indiplon is a highly specific positive allosteric modulator for GABA_A receptors containing the $\alpha 1$ subunit^{27,28}, enriched in PV basket cells synapses²⁹. Indeed, indiplon application *in vitro* (300 nM) enhanced spontaneous IPSCs recorded from layer 2/3 cortical pyramidal cells of *Cntnap4* mutant mice (Extended Data Fig. 10c). Acute administration of indiplon by oral gavage²⁷ restored the startle response in knockout and heterozygotes to wild-type levels (Fig. 4e). At the same dosage, wild-type animal startle was unaffected (Fig. 4e). Indiplon did not however affect PPI in knockout, heterozygote or wild-type animals (Extended Data Fig. 10b). This suggests that the GABAergic system is involved in the proper maturation of sensory-motor processing. Consistent with defective PV cell output, *Cntnap4* mutant mice exhibited mild epileptiform-like activity under deep anaesthesia that was not seen in controls (Extended Data Fig. 3b).

Our results show that *Cntnap4* is located presynaptically and indicate that it functions in two distinct ways depending on the system. In dopaminergic synapses, which work by volume transmission on a relatively

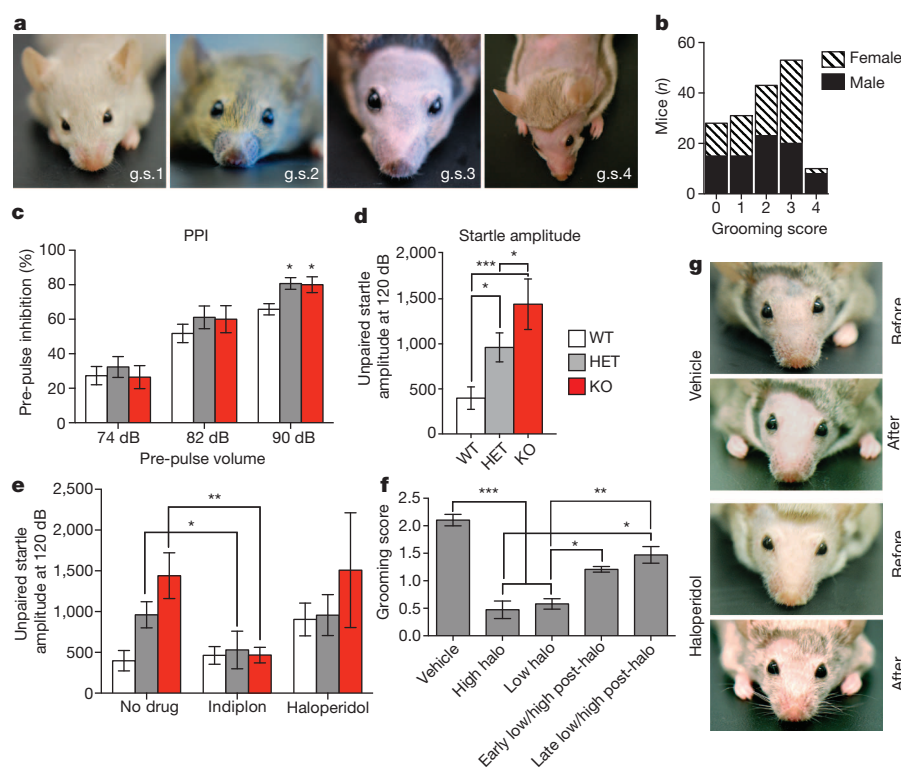


Figure 4 | Aberrant behaviour exhibited by *Cntnap4* mutant mice can be rescued by specific pharmacological intervention. **a**, Over-grooming score (g.s.) in *Cntnap4* mutant mice from 1–4 (1, lacks whiskers; 2, lacks whiskers and snout hair; 3, complete facial hair loss; 4, additional body hair loss). **b**, Grooming score for 165 *Cntnap4* mice. **c**, Increase in pre-pulse inhibition at 90 dB between mutant and wild type (WT, $n = 15$; HET, $n = 12$; KO, $n = 10$; WT versus HET, $P < 0.05$; WT versus KO, $P < 0.05$; one-way ANOVA, Tukey's post-hoc test). **d**, Increase in startle response amplitude in mutant versus wild type (WT, $n = 15$; HET, $n = 12$; KO, $n = 10$; WT versus HET, $P < 0.05$; WT versus KO, $P < 0.001$; HET versus KO, $P < 0.05$; one-way ANOVA, Tukey's post-hoc test). **e**, Startle amplitude in mutant mice was normalized upon acute administration of indiplon, but was unaffected by chronic delivery of haloperidol. **f**, Amelioration of over-grooming behaviour by chronic haloperidol (halo) administration. After the end of the haloperidol regimen, over-grooming returns. Early haloperidol treatment is 90 days after implantation (vehicle, $n = 58$ pups; high haloperidol, $n = 19$ pups; low haloperidol, $n = 44$; early post-haloperidol, $n = 24$ pups; late post-haloperidol, $n = 17$ pups). **g**, Representative vehicle- and haloperidol-treated adult heterozygote before and after 60 days of haloperidol treatment. Re-emergence of facial hair and whiskers in treated mouse only.

slow timescale, mis-regulated release has a prominent role. In contrast, the GABAergic system works through fast synaptic transmission to deliver properly timed inhibition, and as a result the structural abnormality in the synapses predominates and leads to less efficient output with slower kinetics. Thus, in each instance, the kinetics of the system dictates the outcome of the loss of protein. Intriguingly, our findings that the loss of *Cntnap4* results in opposing failures in neurotransmission in the dopaminergic and GABAergic systems are consistent with the deficiencies observed in neuropsychiatric patients^{5,30}.

METHODS SUMMARY

Cntnap4 knock-in mice used in this study were generated by introducing enhanced GFP (eGFP) in frame with *Cntnap4* start codon. To characterize gene expression, eGFP was used to represent *Cntnap4*-expressing cells in the brain. Western blot analysis for *Cntnap4* localization was performed on adult mouse cortex using a *Cntnap4* rabbit polyclonal antibody described previously⁸. The *Cntnap4*-Fc fusion protein was applied to live neuronal cultures for *in vitro* localization studies. Evoked extracellular dopamine release was measured using fast-scan cyclic voltammetry of P60 *in vitro* slice preparations in wild type, heterozygous and knockout animals¹¹. For paired cell recordings between FS and excitatory neurons, *Cntnap4* mice were crossed to a parvalbumin-cre:RFP reporter background to enable targeted physiological recordings. For synaptic ultrastructural analysis, P60 *Cntnap4* wild type, heterozygous and knockout animals were analysed by electron microscopy, focusing on perisomatic inhibitory synaptic contacts. *Cntnap4* mice were assessed by a number of behavioural analyses including grooming score and pre-pulse inhibition^{28,29} (San Diego Instruments). Pharmacological rescue was performed using indiplon (Tocris), which was administered acutely by oral gavage as well as haloperidol (Innovative Research of America), which was delivered chronically via subcutaneous slow-release pellet²⁷.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 January; accepted 11 March 2014.

Published online 25 May; corrected online 9 July 2014 (see full-text HTML version for details).

- Peñagariño, O. *et al.* Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell* **147**, 235–246 (2011).
- Etherton, M. R., Blaiss, C. A., Powell, C. M. & Sudhof, T. C. Mouse neuroligin-1 α deletion causes correlated electrophysiological and behavioral changes consistent with cognitive impairments. *Proc. Natl Acad. Sci. USA* **106**, 17998–18003 (2009).
- Fazzari, P. *et al.* Control of cortical GABA circuitry development by Nrg1 and ErbB4 signalling. *Nature* **464**, 1376–1380 (2010).
- Berghuis, P. *et al.* Brain-derived neurotrophic factor controls functional differentiation and microcircuit formation of selectively isolated fast-spiking GABAergic interneurons. *Eur. J. Neurosci.* **20**, 1290–1306 (2004).
- Lewis, D. A. Cortical circuit dysfunction and cognitive deficits in schizophrenia - implications for preemptive interventions. *Eur. J. Neurosci.* **35**, 1871–1878 (2012).
- Blatt, G. J. & Fatemi, S. H. Alterations in GABAergic biomarkers in the autism brain: research findings and clinical implications. *Anat. Rec.* **294**, 1646–1652 (2011).
- Ashrafi, S. *et al.* Neuronal Ig/Caspr recognition promotes the formation of axoaxonic synapses in mouse spinal cord. *Neuron* **81**, 120–129 (2014).
- Spiegel, I., Salomon, D., Erne, B., Schaefer-Wiemers, N. & Peles, E. Caspr3 and Caspr4, two novel members of the Caspr family are expressed in the nervous system and interact with PDZ domains. *Mol. Cell. Neurosci.* **20**, 283–297 (2002).
- Ho, A., Morishita, W., Hammer, R. E., Malenka, R. C. & Sudhof, T. C. A role for Mints in transmitter release: Mint 1 knockout mice exhibit impaired GABAergic synaptic transmission. *Proc. Natl Acad. Sci. USA* **100**, 1409–1414 (2003).
- Atasoy, D. *et al.* Deletion of CASK in mice is lethal and impairs synaptic function. *Proc. Natl Acad. Sci. USA* **104**, 2525–2530 (2007).
- Patel, J. C. & Rice, M. E. Monitoring axonal and somatodendritic dopamine release using fast-scan cyclic voltammetry in brain slices. *Methods Mol. Biol.* **964**, 243–273 (2013).
- Li, X. *et al.* Enhanced striatal dopamine transmission and motor performance with LRRK2 overexpression in mice is eliminated by familial Parkinson's disease mutation G2019S. *J. Neurosci.* **30**, 1788–1797 (2010).
- Doischer, D. *et al.* Postnatal differentiation of basket cells from slow to fast signaling devices. *J. Neurosci.* **28**, 12956–12968 (2008).
- Hefft, S. & Jonas, P. Asynchronous GABA release generates long-lasting inhibition at a hippocampal interneuron-principal neuron synapse. *Nature Neurosci.* **8**, 1319–1328 (2005).
- Labasque, M. & Faivre-Sarrailh, C. GPI-anchored proteins at the node of Ranvier. *FEBS Lett.* **584**, 1787–1792 (2010).
- Krueger, D. D., Tuffy, L. P., Papadopoulos, T. & Brose, N. The role of neuroligins in the formation, maturation, and function of vertebrate synapses. *Curr. Opin. Neurobiol.* **22**, 412–422 (2012).
- Chang, M. C. *et al.* Narp regulates homeostatic scaling of excitatory synapses on parvalbumin-expressing interneurons. *Nature Neurosci.* **13**, 1090–1097 (2010).
- Sylwestrak, E. L. & Ghosh, A. Elfn1 regulates target-specific release probability at CA1-interneuron synapses. *Science* **338**, 536–540 (2012).
- Cathala, L., Holderith, N. B., Nusser, Z., DiGregorio, D. A. & Cull-Candy, S. G. Changes in synaptic structure underlie the developmental speeding of AMPA receptor-mediated EPSCs. *Nature Neurosci.* **8**, 1310–1318 (2005).
- Peça, J. *et al.* *Shank3* mutant mice display autistic-like behaviours and striatal dysfunction. *Nature* **472**, 437–442 (2011).
- Chao, H. T. *et al.* Dysfunction in GABA signalling mediates autism-like stereotypies and Rett syndrome phenotypes. *Nature* **468**, 263–269 (2010).
- Geyer, M. A., Krebs-Thomson, K., Braff, D. L. & Swerdlow, N. R. Pharmacological studies of prepulse inhibition models of sensorimotor gating deficits in schizophrenia: a decade in review. *Psychopharmacol.* **156**, 117–154 (2001).
- Stark, K. L., Burt, R. A., Gogos, J. A. & Karayiorgou, M. Analysis of prepulse inhibition in mouse lines overexpressing 22q11.2 orthologues. *Int. J. Neuropsychopharmacol.* **12**, 983–989 (2009).
- Champagne, F. A. *et al.* Variations in nucleus accumbens dopamine associated with individual differences in maternal behavior in the rat. *J. Neurosci.* **24**, 4113–4123 (2004).
- Grados, M. A. The genetics of obsessive-compulsive disorder and Tourette's syndrome: what are the common factors? *Curr. Psychiatry Rep.* **11**, 162–166 (2009).
- Crowley, J. J. *et al.* Antipsychotic-induced vacuous chewing movements and extrapyramidal side effects are highly heritable in mice. *Pharmacogenomics J.* **12**, 147–155 (2012).
- Foster, A. C. *et al.* *In vivo* pharmacological characterization of indiplon, a novel pyrazolopyrimidine sedative-hypnotic. *J. Pharmacol. Exp. Ther.* **311**, 547–559 (2004).
- Petroski, R. E. *et al.* Indiplon is a high-affinity positive allosteric modulator with selectivity for $\alpha 1$ subunit-containing GABA_A receptors. *J. Pharmacol. Exp. Ther.* **317**, 369–377 (2006).
- Klausberger, T., Roberts, J. D. B. & Somogyi, P. Cell type- and input-specific differences in the number and subtypes of synaptic GABA_A receptors in the hippocampus. *J. Neurosci.* **22**, 2513–2521 (2002).
- Castellanos, F. X. *et al.* Sensorimotor gating in boys with Tourette's syndrome and ADHD: preliminary results. *Biol. Psychiatry* **39**, 33–41 (1996).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors are grateful to R. Froemke for critically reading the manuscript, to B. Benedetti, M. McKenzie Chang, L. Cobbs, B. A. Heller, T. Petros and N. Yumoto (all NYU) for help with experiments and analysis and to Charles Nicholson (NYU) for providing specialized software to analyse Vmax. Research in the Fishell laboratory is supported by the NIH (grants R01 NS081297, R01 MH071679, R01 NS074972, P01 NS074972 to B.R. and G.F.) and the Simons Foundation (94534). The Rice laboratory is supported by the NIH (grants R01 NS036362 and R01 DA033811) and the Attilio and Olympia Ricciardi Research Fund. The Rudy laboratory is supported by the NIH (NS30989). The Peles laboratory is supported by the NIH (grant NS020220) and the Israel Science Foundation. T.K. support was provided through postdoctoral fellowships from the Patterson Trust and Roche. E.A. support was provided by New York State through its NYSTEM initiative (C024326) and fellowship from Canadian Institutes of Health Research. J.C.P. support was provided by NYU COE Addiction Seed Grant. This work was funded by the Institut Pasteur, INSERM, AP-HP, University Paris Diderot and the Bettencourt-Schueller, Orange, FondaMental, Conny-Maeva, Cognacq-Jay foundations.

Author Contributions T.K., E.A. and G.F. designed the study and wrote the manuscript. T.K. and E.A. performed all the experiments and analysis except for the following: J.C.P. performed the *in vitro* voltammetry experiments and analysed the data. I.K. performed the majority of the *in vitro* paired recordings. M.K., L.R.-M. and S.M. provided the SNP data. J.G. and H.H. provided the intronic CNV data. D.H., B.K., G.H., R.D. and T.B. provided the exonic CNV data. S.R. performed the synaptosome preparation and western blots. D.S. and E.P. made the *Cntnap4* mouse. A.G. performed the verification of the *Cntnap4* mouse. N.C.R. performed the *in vivo* electrophysiological experiments. C.H. set up and advised on behavioural experiments. J.A.G., R.W.T., B.R. and M.E.R. advised on experiments and manuscript preparation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.F. (gordon.fishell@med.nyu.edu).

Novel somatic and germline mutations in intracranial germ cell tumours

Linghua Wang¹, Shigeru Yamaguchi², Matthew D. Burstein^{3,4}, Keita Terashima^{2,5}, Kyle Chang¹, Ho-Keung Ng⁶, Hideo Nakamura⁷, Zongxiao He⁸, Harshavardhan Doddapaneni¹, Lora Lewis¹, Mark Wang¹, Tomonari Suzuki⁹, Ryo Nishikawa⁹, Atsushi Natsume¹⁰, Shunsuke Terasaka¹¹, Robert Dauser¹², William Whitehead¹², Adesina Adekunle¹³, Jiayi Sun⁴, Yi Qiao¹⁴, Gábor Marth¹⁴, Donna M. Muzny¹, Richard A. Gibbs¹, Suzanne M. Leal⁸, David A. Wheeler¹ & Ching C. Lau^{2,4,15}

Intracranial germ cell tumours (IGCTs) are a group of rare heterogeneous brain tumours that are clinically and histologically similar to the more common gonadal GCTs. IGCTs show great variation in their geographical and gender distribution, histological composition and treatment outcomes. The incidence of IGCTs is historically five- to eightfold greater in Japan and other East Asian countries than in Western countries¹, with peak incidence near the time of puberty². About half of the tumours are located in the pineal region. The male-to-female incidence ratio is approximately 3–4:1 overall, but is even higher for tumours located in the pineal region³. Owing to the scarcity of tumour specimens available for research, little is currently known about this rare disease. Here we report the analysis of 62 cases by next-generation sequencing, single nucleotide polymorphism array and expression array. We find the KIT/RAS signalling pathway frequently mutated in more than 50% of IGCTs, including novel recurrent somatic mutations in *KIT*, its downstream mediators *KRAS* and *NRAS*, and its negative regulator *CBL*. Novel somatic alterations in the AKT/mTOR pathway included copy number gains of the *AKT1* locus at 14q32.33 in 19% of patients, with corresponding upregulation of *AKT1* expression. We identified loss-of-function mutations in *BCORL1*, a transcriptional co-repressor and tumour suppressor. We report significant enrichment of novel and rare germline variants in *JMJD1C*, which codes for a histone demethylase and is a coactivator of the androgen receptor, among Japanese IGCT patients. This study establishes a molecular foundation for understanding the biology of IGCTs and suggests potentially promising therapeutic strategies focusing on the inhibition of KIT/RAS activation and the AKT1/mTOR pathway.

IGCTs are divided into two main groups, pure germinoma and non-germinomatous germ cell tumours (NGGCTs). Germinoma is the most common subtype¹, accounting for over two thirds of all IGCTs^{1,4}. NGGCTs include teratoma, embryonal carcinoma, yolk sac tumour and choriocarcinoma. About 10% of germinomas and most NGGCTs remain refractory to multimodality therapy⁴. The biology of these tumours is currently largely unknown, except for gain-of-function mutations of *KIT* reported in ~25% of pure germinomas^{5,6} and a few chromosomal abnormalities revealed by comparative genomic hybridization (CGH)^{7–9}. Through an international multicentre collaboration, we have conducted an in-depth analysis of the genetic abnormalities of IGCTs. For the discovery study, whole-exome sequencing of 28 cases yielded an average of 139× coverage with 95.4% of targeted bases covered by ≥ 20× (Extended Data Fig. 1). We validated a mean of six non-silent mutations per sample (Supplementary Table 1), corresponding to approximately

0.50 non-silent mutations per megabase (Mb, Extended Data Fig. 2a, b). Although there was no significant difference in average mutation rate between pure germinomas and NGGCTs, the mutation rates varied markedly among NGGCTs (Extended Data Fig. 2c). For the validation study, we performed targeted deep sequencing, with an average depth of coverage of about 1,000×, for an additional 34 IGCT cases using a custom-designed AmpliSeq array (Methods). The mutations identified are listed in Supplementary Table 2.

The recurrent genetic alterations and clinical data are summarized in Figs 1 and 2. Except for *KIT*, none of these genes were previously linked to IGCTs. Overall, 53% of the tumours harboured somatic mutations in at least one of the genes involved in KIT/RAS or AKT/mTOR pathways (Fig. 3a, b). 93% of somatic mutations identified in these pathways were predicted to be deleterious (Supplementary Table 3).

Oncogenic *KIT* mutations are common in testicular seminomas¹⁰ and gastrointestinal stromal tumours (GISTs)¹¹, resulting in ligand-independent kinase activation¹². *KIT* was mutated in 16 IGCT tumours (Fig. 2a), but not in any NGGCT cases. Mutations in *KIT* were clustered primarily in exon 17, followed by exon 11 in a pattern similar to that of testicular seminomas¹⁰, but quite different from GISTs¹¹ in which mutations cluster in exon 11, followed by the extracellular domain (Fig. 2a). *KIT* was overexpressed in the majority of pure germinomas (Extended Data Fig. 3), but rarely in NGGCTs. So far, eight tyrosine kinase inhibitors (TKIs) targeting *KIT* have been approved (Supplementary Table 4). Since the long-term side effects of radiation are well recognized, the application of TKIs could potentially benefit IGCT patients through dose reduction or elimination of radiation therapy.

KRAS or *NRAS* were mutated in 19% of IGCT cases (Extended Data Fig. 4a). The *KRAS*/*NRAS* and *KIT* mutations were mutually exclusive genetic events in IGCTs (Fig. 3a, Extended Data Fig. 4b, $P = 0.018$). So far, no approved targeted therapies are available for cancers with *KRAS* mutations. A recent preclinical study of the *MEK* inhibitor, selumetinib, indicated great efficacy against *KRAS*-mutated non-small-cell lung cancer cells¹³. Another study showed effectiveness of *ERK* inhibitors in *MEK* inhibitor-resistant cells¹⁴.

CBL, encoding a RING finger ubiquitin E3 ligase, was the third most frequently mutated gene in IGCTs (Fig. 2b). The protein *CBL* has been shown to function as a negative regulator of receptor protein tyrosine kinases (RPTKs), including *KIT*, by targeting activated receptors for poly-ubiquitination and subsequent degradation¹⁵. *CBL* mutation is associated with *KIT* overexpression, aberrant expression of phosphorylated *STAT5*, and poor prognosis in myeloid malignancies¹⁶. In this study, three somatic mutations were identified in the RING finger domain, which might

¹Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA. ²Texas Children's Cancer and Hematology Centers, Baylor College of Medicine, Houston, Texas 77030, USA. ³Structural and Computational Biology and Molecular Biophysics Program, Baylor College of Medicine, Houston, Texas 77030, USA. ⁴Medical Scientist Training Program, Baylor College of Medicine, Houston, Texas 77030, USA. ⁵National Center for Child Health and Development, Tokyo, 157-8535, Japan. ⁶Department of Anatomical and Cellular Pathology, Prince of Wales Hospital, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong. ⁷Department of Neurosurgery, Kumamoto University, Kumamoto, 860-0862, Japan. ⁸Center for Statistical Genetics, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA. ⁹Department of Neurosurgery, Saitama Medical University, Saitama, 350-0495, Japan. ¹⁰Department of Neurosurgery, Nagoya University, Nagoya, 466-8550, Japan. ¹¹Department of Neurosurgery, Hokkaido University, Hokkaido Prefecture, 060-0808, Japan. ¹²Department of Neurosurgery, Baylor College of Medicine, Houston, Texas 77030, USA. ¹³Department of Pathology, Baylor College of Medicine, Houston, Texas 77030, USA. ¹⁴Department of Biology, Boston College, Chestnut Hill, Maryland 02467, USA. ¹⁵Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, Texas 77030, USA.

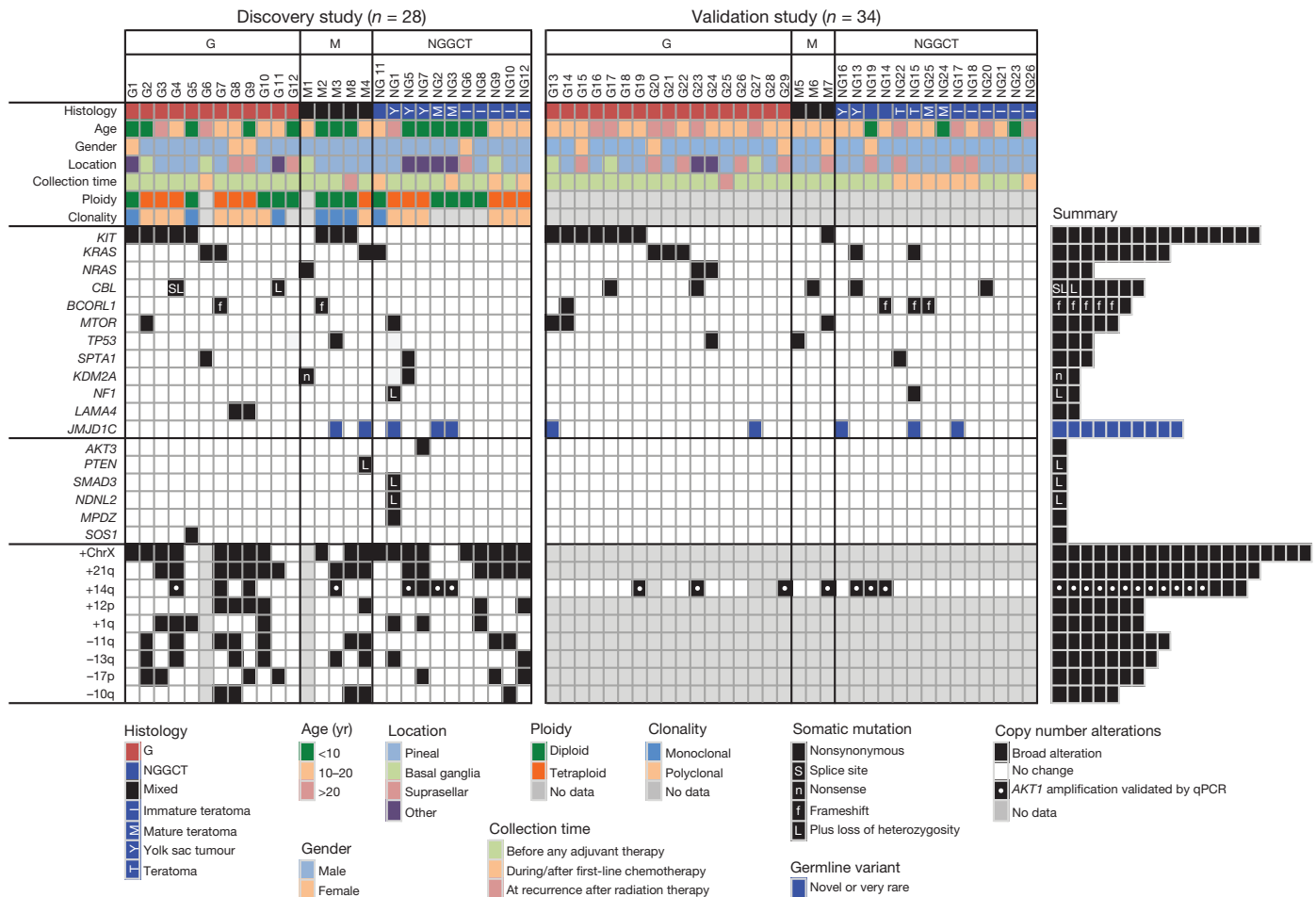


Figure 1 | Subgroup specificity of the recurrent genetic alterations identified in 62 IGCT patients. G, germinoma; M, mixed germ cell tumour with germinoma component; NGGCT, non-germinomatous germ cell tumour. Several genes were included that are mutated once but considered biologically important by one of the following criteria: involvement in KIT/RAS or

abolish *CBL*-directed polyubiquitination and downregulation of RPTKs¹⁵, and two recurrent mutations were found in the small linker domain, which are supposed to be oncogenic¹⁵ (Fig. 2b). *KIT* was overexpressed in tumour G23 that was wild type for *KIT* but had a somatic mutation in *CBL* (Extended Data Fig. 3), consistent with its role as a negative regulator of *KIT*. A recent report showed that dasatinib is the most effective TKI against myeloid leukaemia cell lines that harbour homozygous *CBL* mutations¹⁷, and therefore might be an effective agent for the treatment of *CBL*-mutated IGCTs. We further analysed the copy number status of *CBL* and found that 13 out of 28 cases had either clonal or subclonal 11q loss of heterozygosity (LOH) spanning the *CBL* locus (Supplementary Table 5 and Extended Data Fig. 5a). Tumours G4 and G11 with *CBL* somatic mutation lost the wild-type allele through clonal and subclonal 11qLOH, respectively (Extended Data Fig. 5b), further reinforcing the notion that *CBL* might have an important role in the pathogenesis of IGCTs.

We observed focal amplification of 14q32.33 in five tumours (18%) in the discovery set (Extended Data Fig. 6a). The amplified region included 2 Mb spanning the *AKT1* locus. Among 34 genes within this region, *AKT1* was the only known oncogene. We validated *AKT1* copy number gains by quantitative PCR for all five cases and seven additional cases in the validation set. Copy number gains at 14q32.33 were associated with elevated messenger RNA expression of *AKT1* (Fig. 3c), but not for other genes in this region (Supplementary Table 6 and Extended Data Fig. 6b). We also observed strong cytoplasmic and nuclear immunostaining of AKT1 in cases with focal *AKT1* copy number gains (Fig. 3d).

AKT/mTOR pathways, known interaction with *KIT*, tumour suppressor genes with two hits. Germline variants in *JMJD1C* were either novel or rare polymorphisms with minor allele frequency less than 0.005. qPCR of *AKT1* was validated if the gene count adjusted by ploidy was greater than 3.

In this study, 75% cases with *AKT1* copy number gains were identified in tumours with wild-type *KIT*, *KRAS* and *NRAS* (Fig. 3a), indicating an important role of AKT/mTOR pathway in IGCTs. The serine-threonine protein kinase AKT is a key intermediate of signalling pathways that regulate cellular processes. AKT activation by mutation is frequently seen in human cancers, but overexpression is rarely reported. However, high-level *AKT1* expression was reported in non-small-cell lung cancer, where it was correlated with poor response to cisplatin¹⁸. Use of *AKT1* inhibitors might also be a promising strategy for treating IGCT patients with *AKT1* overexpression.

Finally, we identified recurrent somatic mutations in *BCORL1*, *MTOR*, *TP53*, *SPTA1*, *KDM2A* and *LAMA4* (Fig. 1). *BCORL1*, a transcriptional co-repressor located on the X-chromosome (Xq25-q26.1), was thought to be a tumour suppressor gene in acute myelogenous leukaemia¹⁹. Functional studies have shown that *BCORL1* can interact with class II histone deacetylases, the CtBP co-repressor, through the CtBP binding motif-PLDLS and affect the repression of E-cadherin²⁰. In this study, *BCORL1* was mutated in six male patients (Fig. 2b). Interestingly, in five of them, mutations were out-of-frame insertions/deletions, potentially resulting in a truncated protein lacking the conserved LXXLL motif, which mediates the hormone-induced interaction between nuclear receptor and co-activators²¹.

Given the low mutation rate and early onset of this disease, we sought to identify potential IGCT-predisposing germline mutations. Therefore we screened germline sequence data for genes enriched in novel functional variants in the IGCT discovery cohort compared to control

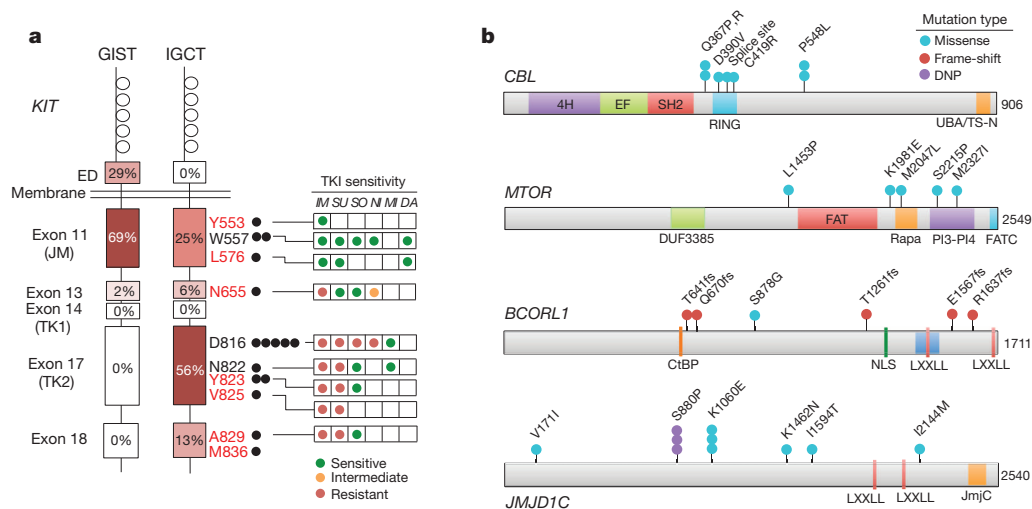


Figure 2 | Novel recurrent somatic and germline mutations in IGCT.

a, Somatic *KIT* mutations. Red lettering, the novel *KIT* mutations identified in IGCT; black lettering, reported *KIT* mutations; black-filled circles, the number of mutations identified at each mutation site. The primary *KIT* mutations reported in gastrointestinal stromal tumours are shown for comparison. Functional domains are ED, extracellular domain; JM, juxtamembrane domain; TK1, tyrosine kinase I, the ATP binding domain; TK2, tyrosine kinase II, the kinase activation loop (A-loop). The sensitivity of known TKIs, corresponding to each mutation site from previous studies performed in other tumour types, are shown on the right. IM, imatinib; SU, sunitinib, SO, sorafenib; NI, nilotinib; MI, midostaurin; DA, dasatinib. *KIT* exon-11 mutations are generally sensitive to imatinib. Certain imatinib-resistant *KIT*

mutations respond to sunitinib and sorafenib. More than half of *KIT* mutations in IGCTs reside in the A-loop (Supplementary Fig. 3). The D816 mutation causes *KIT* to be constitutively activated by altering the structure of the JM domain and destabilizing the A-loop inactive conformation¹². Tumours with D816-mutated *KIT* respond well to midostaurin. **b**, Schematic representation of somatic mutations in *CBL*, *MTOR*, *BCORL1*, and germline variants identified in *JMJD1C*. DNP, dinucleotide polymorphism; Rapa, rapamycin binding site; PI3-PI4, PI3-PI4 kinase; CtBP, CtBP binding site; NLS, nuclear localization signal; LXXLL motif, L is leucine and X is any amino acid. Only novel or rare *JMJD1C* germline variants with minor allele frequency (MAF) < 0.005 are shown.

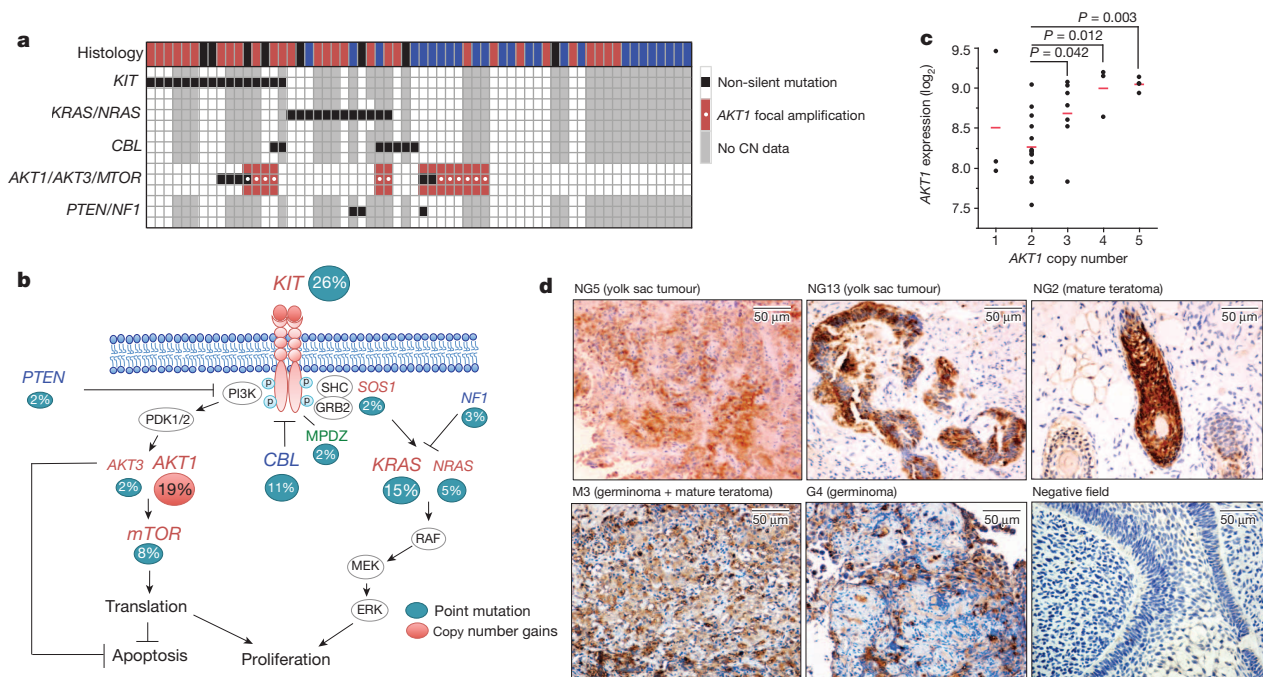


Figure 3 | Frequent genetic alterations of the KIT/RAS and AKT/mTOR signalling pathways. **a**, Summary of the somatic events. CN, copy number. **b**, KIT/RAS and AKT/mTOR pathway interactions showing frequencies of somatic alterations in key genes. Alteration frequencies are expressed as a percentage of all IGCT patients. Red text, protein positively regulates signalling; blue text, protein negatively regulates signalling; green text, physically interacting protein. **c**, the correlation of *AKT1* copy number status and levels of *AKT1* mRNA expression. *AKT1* copy number status was assayed by single nucleotide polymorphism array and validated by qPCR. The mRNA expression levels were determined by Affymetrix U133Plus2 human gene expression array.

The red lines indicate mean values of expression. The P value across all groups calculated by Spearman's rank-order correlation analysis is 0.001 and the correlation coefficient is 0.5614. The P values between two different groups calculated by one-way ANOVA analysis are shown. **d**, Immunohistochemical staining of AKT1 in AKT1 amplified tumours. Immunostaining was carried out with AKT1-specific goat antibody D-17 (sc7126, Santa-Cruz Biotechnology) at 1:75 dilution. Magnification, $\times 400$. Scale bars, 50 μ m. Cases M3 and NG5 showed strong and diffuse cytoplasmic and nuclear staining, whereas cases G4, NG2 and NG13 showed strong but focal cytoplasmic and nuclear staining.

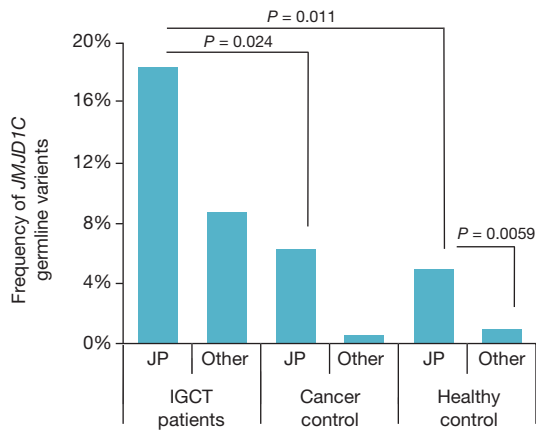


Figure 4 | Enrichment of germline *JMJD1C* variants in IGCT. Only validated novel or rare (MAF < 0.005) non-silent germline variants were counted. JP, Japanese. Healthy controls were from the 1000 Genomes Project ($n = 1,092$, <http://www.1000genomes.org/>). A panel of 778 unrelated cancers sequenced at HGSC (cancer types = 10, from both Asia and United States) were used as the cancer control. The one-sided Fisher's exact test was performed to test the enrichment of the minor alleles in IGCT cases.

populations (Extended Data Fig. 7 and Supplementary Table 7). The top four genes most enriched in functional germline variants were *CDK5RAP2*, *JMJD1C*, *USP35* and *PCDH15*. Of these, *JMJD1C* functions as a chromatin modifier gene, and was the only one with a reported role in normal germinal tissue development in both human²² and mouse models²³. A recent report showed that it is required in the development and maintenance of germ cells in mice²⁴. Thus we included this gene for mutation screening in the validation cohort. In the combined cohorts of 62 cases, we identified novel and rare germline variants in *JMJD1C* in 10 patients, including a rare dinucleotide polymorphism (AA to GC, S880P) in three genetically unrelated individuals (Fig. 2b and Supplementary Table 8). Among the 10 patients, 9 carriers were from Japan and 1 was from Hong-Kong. *JMJD1C* germline variants were significantly enriched in the Japanese populations in control cohorts and further enriched (about fivefold) in Japanese IGCT patients (Fig. 4). The odds ratio is 4.8, indicating a strong association between *JMJD1C* variants and the risk of developing IGCT. The rare variant association tests^{25,26} also revealed significant association of *JMJD1C* genotype with this disease (Supplementary Table 9). Like *BCORL1*, there are two conserved LXXLL nuclear-receptor-interacting motifs at the carboxy-terminus of *JMJD1C*, which mediate the interaction with nuclear receptors. Indeed, *JMJD1C* is known to interact with the thyroid hormone receptor in the thyroid²⁷ and was reported to interact with the androgen receptor (AR) in humans²⁸. Expression microarrays revealed high *JMJD1C* and AR expression levels in all 37 IGCT patients tested (Extended Data Fig. 8).

DNA copy-number and clonality analysis (Methods) of IGCT tumours using single nucleotide polymorphism array data revealed a heterogeneous and complex tumour genome. Nearly half of the tumours were tetraploid and 71% of the analysed cases showed subclonal structure (Extended Data Figs 9 and 10). We observed almost exclusively chromosomal arm-level gains and losses. Chromosome X, 21q, 12p, 1q and 14q were frequently amplified, whereas 11q, 10q, 17p and 13q were frequently deleted. Only copy number gains of X, 12p and loss of 13q have been described previously²⁹. Over 90% of the pure germinomas exhibited chromosomal imbalance and acquired uniparental disomy of multiple chromosomes, whereas 40% of the NGGCTs were chromosomally stable. The preponderance of arm-level and chromosome-level imbalance suggests the possibility that aberrant meiotic division may have a role in the pathogenesis of pure germinomas.

Although pure germinomas respond well to surgery and radiotherapy, the long-term quality of life for about a third of postoperative patients is severely impaired owing to radiotherapy-associated complications.

In addition, nearly 10% of patients develop recurrence even after multi-modality therapy^{1,4}. Except for mature teratomas, most NGGCTs remain refractory to available treatments and have a rather poor prognosis^{1,4}. In this study we identified novel and frequent somatic alterations in the KIT/RAS and AKT/mTOR signalling pathways and suggested potentially promising targets for the development of novel therapeutic strategies focusing on inhibition of these pathways. Although fewer therapeutic targets were found in NGGCTs, frequent *AKT1* amplification and recurrent mTOR mutations may open the door to possible use of AKT1/mTOR inhibitors. It is intriguing that both histone-modifying genes discovered in this study are implicated in the interaction with nuclear receptor proteins. It suggests the possibility that *JMJD1C* and *BCORL1* might be associated with the male preponderance and age of peak incidence of IGCTs through interaction with AR triggered by elevated levels of androgen hormones at puberty. Further functional studies using mouse models are needed to elucidate the possible role of *JMJD1C* in the pathogenesis of this disease. This study comprehensively described the genomic heterogeneity and complexity underlying IGCTs and provides a molecular foundation for understanding of the genetic alterations in IGCT.

METHODS SUMMARY

Genomic DNA was extracted following the manufacturer's protocols. Whole-exome sequencing was carried out with NimbleGen in-solution reagents (VCRome2.1, NimbleGen). The target region is 43 Mb. Enriched libraries were sequenced on an Illumina HiSeq2000 platform (Illumina). Whole-exome sequencing data processing and analysis were performed using the standard pipelines (Methods) established at the Human Genome Sequencing Center (HGSC), Baylor College of Medicine. Mutation validation was performed using the Ion Torrent Personal Genome Machine (PGM, Life Technologies). The average coverage was 2,533× and 1,923× for tumour and their matched blood pools, respectively. The validation rates were 93% and 89% for somatic substitutions and indels, respectively. A subset of germline variants was validated using Sanger sequencing. For prevalence screening, 23 genes were selected based on the discovery study results (Methods) and an AmpliSeq custom array was designed (Supplementary Table 10). The AmpliSeq libraries were constructed from 20 ng DNA (10 ng for each reaction pool) and deep sequencing was performed using Ion Torrent PGM platform. Ion Torrent sequencing data was analysed using Torrent Suite Software v3.0. The average coverage was 891×. DNA copy number analyses were performed using Illumina HumanOmni2.5-8 BeadChip Kit. Single nucleotide polymorphism calls were made using Illumina GenomeStudio Version 2011.1 software. The allele-specific copy number, ploidy and purity were analysed by Genome Alteration Print (GAP)³⁰. The focal DNA copy number changes and the clonality analysis were performed using in-house algorithms. Focal *AKT1* amplification was validated by quantitative PCR. Immunostaining was carried out with AKT1-specific goat antibody D-17 (sc-7126, Santa-Cruz Biotechnology) at 1:75 dilution. A complete description of the materials and methods is provided in Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 June 2013; accepted 28 March 2014.

Published online 4 June 2014.

- Packer, R. J., Cohen, B. H. & Cooney, K. Intracranial germ cell tumors. *Oncologist* **5**, 312–320 (2000).
- Jennings, M. T., Gelman, R. & Hochberg, F. Intracranial germ-cell tumors: natural history and pathogenesis. *J. Neurosurg.* **63**, 155–167 (1985).
- McCarthy, B. J. *et al.* Primary CNS germ cell tumors in Japan and the United States: an analysis of 4 tumor registries. *Neuro Oncol.* **14**, 1194–1200 (2012).
- Matsutani, M. *et al.* Primary intracranial germ cell tumors: a clinical analysis of 153 histologically verified cases. *J. Neurosurg.* **86**, 446–455 (1997).
- Kamakura, Y., Hasegawa, M., Minamoto, T., Yamashita, J. & Fujisawa, H. *c-kit* gene mutation: common and widely distributed in intracranial germinomas. *J. Neurosurg.* **104**, 173–180 (2006).
- Sakuma, Y. *et al.* *c-kit* gene mutations in intracranial germinomas. *Cancer Sci.* **95**, 716–720 (2004).
- Rickert, C. H., Simon, R., Bergmann, M., Dockhorn-Dworniczak, B. & Paulus, W. Comparative genomic hybridization in pineal germ cell tumors. *J. Neuropathol. Exp. Neurol.* **59**, 815–821 (2000).
- Schneider, D. T. *et al.* Molecular genetic analysis of central nervous system germ cell tumors with comparative genomic hybridization. *Mod Pathol.* **19**, 864–873 (2006).

9. Terashima, K. *et al.* Genome-wide analysis of DNA copy number alterations and loss of heterozygosity in intracranial germ cell tumors. *Pediatr. Blood Cancer* **61**, 593–600 (2014).
10. Kemmer, K. *et al.* KIT mutations are common in testicular seminomas. *Am. J. Pathol.* **164**, 305–313 (2004).
11. Corless, C. L., Barnett, C. M. & Heinrich, M. C. Gastrointestinal stromal tumours: origin and molecular oncology. *Nature Rev. Cancer* **11**, 865–878 (2011).
12. Gajiwala, K. S. *et al.* KIT kinase mutants show unique mechanisms of drug resistance to imatinib and sunitinib in gastrointestinal stromal tumor patients. *Proc. Natl Acad. Sci. USA* **106**, 1542–1547 (2009).
13. Jänne, P. A. *et al.* Selumetinib plus docetaxel for KRAS-mutant advanced non-small-cell lung cancer: a randomised, multicentre, placebo-controlled, phase 2 study. *Lancet Oncol.* **14**, 38–47 (2013).
14. Hatzivassiliou, G. *et al.* ERK inhibition overcomes acquired resistance to MEK inhibitors. *Mol. Cancer Ther.* **11**, 1143–1154 (2012).
15. Thien, C. B., Walker, F. & Langdon, W. Y. RING finger mutations that abolish c-Cbl-directed polyubiquitination and downregulation of the EGF receptor are insufficient for cell transformation. *Mol. Cell* **7**, 355–365 (2001).
16. Makishima, H. *et al.* Mutations of E3 ubiquitin ligase Cbl/family members constitute a novel common pathogenic lesion in myeloid malignancies. *J. Clin. Oncol.* **27**, 6109–6116 (2009).
17. Makishima, H. *et al.* CBL mutation-related patterns of phosphorylation and sensitivity to tyrosine kinase inhibitors. *Leukemia* **26**, 1547–1554 (2012).
18. Liu, L. Z. *et al.* AKT1 amplification regulates cisplatin resistance in human lung cancer cells through the mammalian target of rapamycin/p70S6K1 pathway. *Cancer Res.* **67**, 6325–6332 (2007).
19. Li, M. *et al.* Somatic mutations in the transcriptional corepressor gene *BCORL1* in adult acute myelogenous leukemia. *Blood* **118**, 5914–5917 (2011).
20. Pagan, J. K. *et al.* A novel corepressor, BCoR-L1, represses transcription through an interaction with CtBP. *J. Biol. Chem.* **282**, 15248–15257 (2007).
21. Heery, D. M., Kalkhoven, E., Hoare, S. & Parker, M. G. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* **387**, 733–736 (1997).
22. Jin, G. *et al.* Genome-wide association study identifies a new locus *JMJD1C* at 10q21 that may influence serum androgen levels in men. *Hum. Mol. Genet.* **21**, 5222–5228 (2012).
23. Kim, S. M. *et al.* Regulation of mouse steroidogenesis by WHISTLE and *JMJD1C* through histone methylation balance. *Nucleic Acids Res.* **38**, 6389–6403 (2010).
24. Kuroki, S. *et al.* *JMJD1C*, a JmjC domain-containing protein, is required for long-term maintenance of male germ cells in mice. *Biol. Reprod.* **89**, 93 (2013).
25. Li, B. & Leal, S. M. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am. J. Hum. Genet.* **83**, 311–321 (2008).
26. Madsen, B. E. & Browning, S. R. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet.* **5**, e1000384 (2009).
27. Katoh, M. & Katoh, M. Identification and characterization of TRIP8 gene in silico. *Int. J. Mol. Med.* **12**, 817–821 (2003).
28. Wolf, S. S., Patchev, V. K. & Obendorf, M. A novel variant of the putative demethylase gene, s-JMJD1C, is a coactivator of the AR. *Arch. Biochem. Biophys.* **460**, 56–66 (2007).
29. Okada, Y., Nishikawa, R., Matsutani, M. & Louis, D. N. Hypomethylated X chromosome gain and rare isochromosome 12p in diverse intracranial germ cell tumors. *J. Neuropathol. Exp. Neurol.* **61**, 531–538 (2002).
30. Popova, T. *et al.* Genome Alteration Print (GAP): a tool to visualize and mine complex cancer genomic profiles obtained by SNP arrays. *Genome Biol.* **10**, R128 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by research funding from the National Human Genome Research Institute (NHGRI, grant number 5U54HG003273) to D.A.W., the Children Brain Tumor Foundation, the Gillson Longenbaugh Foundation and Anderson Charitable Foundation to C.C.L., the CCBTP Fellowship from the Cancer Prevention & Research Institute of Texas (CPRIT grant RP101489) to S.Y., the St Baldrick's Foundation to K.T. and the NLM predoctoral fellowships to M.D.B. (5T15 LM07093-18) and J.S. (5T15 LM07093-19). We thank H. H. Dinh and Y. Han for their technical support, J. G. Reid for Illumina sequence mapping.

Author Contributions L.W. conducted the bioinformatics analyses of the sequencing and single nucleotide polymorphism array data, integrated data from multiple platforms, wrote and revised the manuscript. S.Y. contributed to the conduct of the research. M.D.B., J.S., M.W. contributed to DNA copy number analysis. K.T. contributed to the coordination and conduct of the research. K.C. contributed to the mutation calling and annotation pipeline for AmpliSeq data. H.-K.N. and H.N. performed AKT1 immunohistochemistry assay. L.L. contributed to the construction of the AmpliSeq libraries. C.C.L., T.S., R.N., H.N., A.N., S.T., H.K.N., R.D., W.W. and A.A. collected tumour specimens, provided the histopathological confirmation and interpreted the clinical data. Y.Q., G.M. and L.W. contributed to clonality analysis. D.M.M. and H.D. managed the production pipeline. Z.H. and S.M.L. performed rare variant association tests for *JMJD1C*. R.A.G. contributed to the revision of the manuscript. D.A.W. and C.C.L. conceived the study, supervised the research, and contributed to the writing and revision of the manuscript.

Author Information All sequencing and genotyping data have been deposited in the NCBI database of Genotypes and Phenotypes (dbGaP, <http://www.ncbi.nlm.nih.gov/gap>) under accession number phs000725.v1.p1. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.A.W. (wheeler@bcm.edu) or C.C.L. (cclau@txch.org).

SOX2 controls tumour initiation and cancer stem-cell functions in squamous-cell carcinoma

Soufiane Boumahdi¹, Gregory Driessens^{1*}, Gaëlle Lapouge^{1*}, Sandrine Rorive^{2,3}, Dany Nassar¹, Marie Le Mercier², Benjamin Delatte⁴, Amélie Caauwe¹, Sandrine Lenglez¹, Erwin Nkusi¹, Sylvain Brohée⁵, Isabelle Salmon^{2,3}, Christine Dubois¹, Veronique del Marmol⁶, Francois Fuks⁴, Benjamin Beck¹ & Cédric Blanpain^{1,7}

Cancer stem cells (CSCs) have been reported in various cancers, including in skin squamous-cell carcinoma (SCC)^{1–4}. The molecular mechanisms regulating tumour initiation and stemness are still poorly characterized. Here we find that *Sox2*, a transcription factor expressed in various types of embryonic and adult stem cells^{5,6}, was the most upregulated transcription factor in the CSCs of squamous skin tumours in mice. SOX2 is absent in normal epidermis but begins to be expressed in the vast majority of mouse and human pre-neoplastic skin tumours, and continues to be expressed in a heterogeneous manner in invasive mouse and human SCCs. In contrast to other SCCs, in which SOX2 is frequently genetically amplified⁷, the expression of SOX2 in mouse and human skin SCCs is transcriptionally regulated. Conditional deletion of *Sox2* in the mouse epidermis markedly decreases skin tumour formation after chemical-induced carcinogenesis. Using green fluorescent protein (GFP) as a reporter of *Sox2* transcriptional expression (SOX2–GFP knock-in mice), we showed that SOX2-expressing cells in invasive SCC are greatly enriched in tumour-propagating cells, which further increase upon serial transplantations. Lineage ablation of SOX2-expressing cells within primary benign and malignant SCCs leads to tumour regression, consistent with the critical role of SOX2-expressing cells in tumour maintenance. Conditional *Sox2* deletion in pre-existing skin papilloma and SCC leads to tumour regression and decreases the ability of cancer cells to be propagated upon transplantation into immunodeficient mice, supporting the essential role of SOX2 in regulating CSC functions. Transcriptional profiling of SOX2–GFP-expressing CSCs and of tumour epithelial cells upon *Sox2* deletion uncovered a gene network regulated by SOX2 in primary tumour cells *in vivo*. Chromatin immunoprecipitation identified several direct SOX2 target genes controlling tumour stemness, survival, proliferation, adhesion, invasion and paraneoplastic syndrome. We demonstrate that SOX2, by marking and regulating the functions of skin tumour-initiating cells and CSCs, establishes a continuum between tumour initiation and progression in primary skin tumours.

Skin SCC is the second most frequent skin cancer and affects more than 500,000 new patients per year throughout the world⁸. Recently, different groups have identified CSCs or tumour-propagating cells (TPCs) expressing a high level of CD34 from primary mouse SCC induced by 9,10-dimethyl-1,2-benzanthracene (DMBA)/2-otetradecanoyl phorbol-13-acetate (TPA) treatment that have a higher capacity than other tumour cells to reform secondary tumours upon transplantation into immunodeficient mice and possess higher long-term self-renewal capacity^{1–4}. Very little is known about the mechanisms that regulate cutaneous CSC functions.

To define the intrinsic mechanisms that regulate CSC functions in skin tumours, we determined which transcription factors are differentially expressed between CD34⁺ and CD34[−] Epcam⁺ tumour epithelial

cells (TECs) (Lin[−]Epcam⁺) in benign papillomas and malignant SCCs. *Sox2*, a transcription factor expressed in a great variety of stem cells⁵ and genetically amplified in different cancers, including SCCs from lung and oesophagus⁷, was the most upregulated transcription factor expressed in CD34⁺ TECs from skin papillomas³. Quantitative reverse transcription polymerase chain reaction (qRT–PCR) showed that *Sox2* messenger RNA was indeed highly upregulated in CD34⁺ TECs, and even further upregulated in CD34⁺ TECs of SCCs (Fig. 1a). Using SOX2–GFP knock-in mice⁹, which express GFP under the endogenous regulatory region of *Sox2*, we found that SOX2 expression was absent in the normal epidermis (Fig. 1b), with the exception of in Merkel cells, as previously reported^{10,11}. However, upon drug treatment that induced skin hyperproliferation (retinoic acid, TPA, DMBA or DMBA/TPA), focal expression of SOX2–GFP is detected mostly in suprabasal cells of hyperplastic lesions (Fig. 1b and Extended Data Fig. 1). Nonetheless, no SOX2 protein expression is observed at this stage (Extended Data Fig. 1h), possibly owing to the low level of SOX2 expression. Immunostaining and fluorescence-activated cell sorting (FACS) analysis revealed that SOX2 was expressed in a heterogeneous manner at the mRNA and protein level in more than 80% of primary mouse papillomas and SCCs (Fig. 1c–e and Extended Data Fig. 2a), preferentially but not exclusively expressed in a subset of CD34⁺ TECs (Fig. 1c, d and Extended Data Fig. 2b–d). SOX2 is expressed with a similar frequency and heterogeneity in DMBA/TPA- and Kras(G12D)-induced papilloma and SCCs (Fig. 1e and Extended Data Fig. 2), suggesting that the cellular origin of tumour cells and the mutational heterogeneity are not the primary determinant of SOX2 expression. Immunostaining of a collection of human specimens revealed that SOX2 was expressed in the majority of pre-neoplastic skin lesions—termed actinic keratoses (29/40 patients)—and continued to be expressed in a heterogeneous manner in the majority of invasive SCCs (25/39 patients) (Fig. 1f and Extended Data Fig. 3).

SOX2 is genetically amplified in SCCs from different tissues such as the oesophagus⁷, where SOX2 is naturally expressed⁶. However, no genetic amplification of *Sox2* was detected in mouse and human skin tumour cells (Fig. 1g, h and Extended Data Fig. 4). *Ezh1/2* deletion, which prevents the deposition of H3K27me3 repressive histone marks, increases SOX2 expression in the developing skin epidermis¹². Using chromatin immunoprecipitation (ChIP) experiments, we found that the *Sox2* promoter was associated with repressive histone marks in normal skin and with active histone marks in SCCs (Fig. 1i, j). Altogether, these data indicate that *Sox2* is not genetically amplified but transcriptionally and/or epigenetically regulated during skin tumour initiation in both mice and humans.

To determine whether SOX2 expression during the initial stages of skin tumour initiation is required for tumour formation, we investigated the rate and number of tumours formed upon DMBA/TPA treatment in *Sox2*-deficient epidermis (K14Cre:SOX2^{fl/fl}; ref. 13) (Extended Data

¹Université Libre de Bruxelles, IRIBHM, Brussels B-1070, Belgium. ²Department of Pathology, Erasme Hospital, Université Libre de Bruxelles, Brussels B-1070, Belgium. ³DIAPATH—Center for Microscopy and Molecular Imaging (CMMI), Gosselies B-6041, Belgium. ⁴Laboratory of Cancer Epigenetics, Université Libre de Bruxelles, Brussels B-1070, Belgium. ⁵Machine Learning Group, Computer Science Department, Faculté des Sciences, Université Libre de Bruxelles, Brussels B-1050, Belgium. ⁶Department of Dermatology, Erasme Hospital, Université Libre de Bruxelles, Brussels B-1070, Belgium. ⁷WELBIO, Université Libre de Bruxelles, Brussels B-1070, Belgium.

*These authors contributed equally to this work.

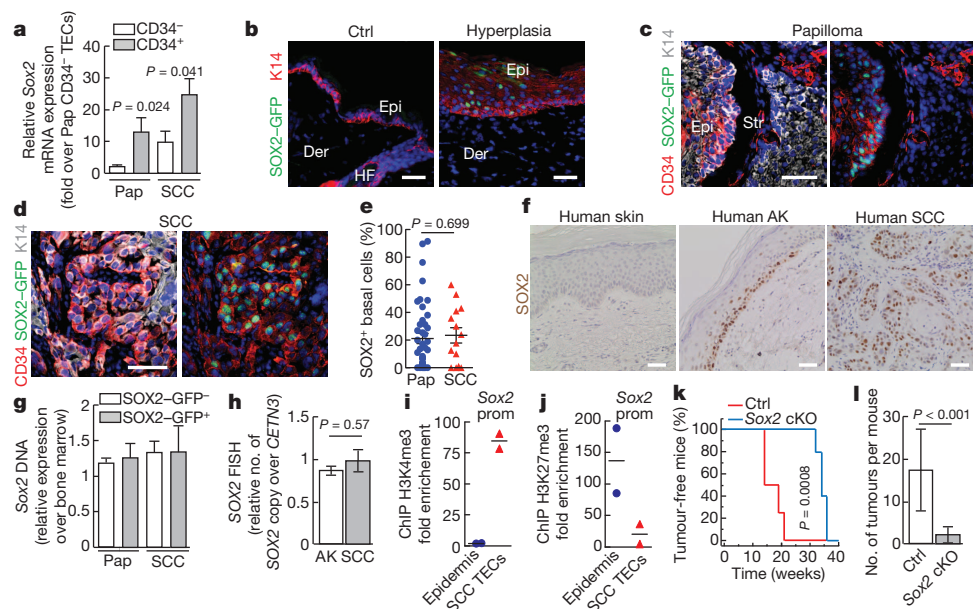


Figure 1 | SOX2 is expressed in pre-neoplastic skin tumours, invasive SCCs and regulates skin tumour initiation. **a**, qRT-PCR analysis of *Sox2* mRNA expression in FACS-isolated CD34⁺ and CD34⁻ TECs ($n = 5$ tumours for each group). Pap, papilloma. **b–d**, Immunostaining of SOX2-GFP, CD34 and K14 showed the presence of SOX2-GFP⁺ cells in hyperplastic skin and in benign and malignant tumours. Ctrl, control; Der, dermis; Epi, epidermis; HF, hair follicle; Str, tumour stroma. **e**, Quantification of the proportion of basal SOX2⁺ TECs within benign and malignant tumours measured by immunofluorescence ($n \geq 1,600$ total cell counted in $n = 14$ tumours from 14 mice). **f**, Immunostaining for SOX2 in human normal skin, actinic keratosis (AK) and invasive SCC. **g**, Quantification of *Sox2* DNA copy number by qRT-PCR of FACS-isolated SOX2-GFP⁺ and SOX2-GFP⁻ TECs. Data were normalized relatively to the house keeping genes *Gabra* and β -actin in tumour cells and in bone-marrow cells (germline DNA) of each mouse ($n = 3$ tumours

from 3 different mice in each group). **h**, Quantification of SOX2 DNA copy number by fluorescence *in situ* hybridization (FISH) in human actinic keratosis and SCC. The data were normalized to the number of CETN3 dots ($n \geq 50$ nuclei counted per AK ($n = 3$) and per SCC ($n = 5$)). **i**, **j**, ChIP of H3K4me3 (**i**) and H3K27me3 (**j**) at the *Sox2* promoter (prom) in normal epidermis and SCC TECs ($n = 2$ skins and $n = 2$ primary SCCs). Fold change represents the percentage of input recovered after immunoprecipitation with the antibody of interest normalized over the percentage of input recovered with beads alone. **k**, Percentage of control and *Sox2* conditional knockout (cKO) mice presenting with skin tumours over time ($n = 5$ mice in each group). **l**, Average number of tumours per mice 40 weeks after tumour initiation (control $n = 7$; *Sox2* conditional knockout $n = 5$ mice). **b**, **c**, **d**, **f**, Hoechst nuclear staining is represented in blue; scale bars, 50 μ m. Data represent the mean and standard error of the mean (s.e.m.).

Fig. 5a, b). Although conditional *Sox2* deletion has no effect on the development and homeostasis of the skin epidermis (Extended Data Fig. 5c, d), it delayed the appearance of skin tumours by several months and decreased the number of tumours by almost tenfold as compared with littermate controls (Fig. 1k, l and Extended Data Fig. 5e, f), indicating that SOX2 has an essential role during the early stages of skin tumour initiation.

At present, the gold standard assay to assess CSC potential is the transplantation of limiting dilutions of highly purified prospectively identified cancer cell populations into immunodeficient mice to assess their ability to form secondary tumours^{14,15}. Transplantation of SOX2-GFP⁺ TECs from invasive SCC (Extended Data Fig. 6a) showed that they had a much greater ability to reform secondary tumours compared with Epcam⁺ SOX2-GFP⁻ and total Epcam⁺ TECs (Fig. 2a, b), and this difference further increased during serial transplantation (Fig. 2c). To assess the respective enrichment of TPCs in cells expressing CD34 and SOX2, we compared the efficiency of secondary tumour formation after the transplantation of all four Epcam⁺ TEC populations (SOX2-GFP⁺/CD34⁺, SOX2-GFP⁺/CD34⁻, SOX2-GFP⁻/CD34⁺ and SOX2-GFP⁻/CD34⁻) (Extended Data Fig. 6b). Interestingly, SOX2-GFP⁺ CD34⁺ double-positive cells showed the greatest ability to reform secondary tumours (Fig. 2d). SOX2-GFP⁺/CD34⁻ TECs, although much less efficient than SOX2-GFP⁺/CD34⁺ TECs, were more efficient than SOX2-GFP⁻ cells irrespective of CD34 expression (Fig. 2d). These results further support the idea that SOX2 is the primary determinant of TPC frequency and identify SOX2⁺/CD34⁺ TECs as the most efficient subpopulation of TPCs in skin SCC.

Grafted SOX2-GFP⁺ TECs reformed tumours that recapitulated the histology of the primary SCCs, with a concomitant increase in the proportion of SOX2-expressing cells with serial transplantation (Extended

Data Fig. 6a, c). The proportion of SOX2-GFP⁺ cells after serial transplantation of all Epcam⁺ TECs increased from 25% in the primary tumours to 80% after the second transplantation (Extended Data Fig. 6d–f), supporting the selective growth advantage of SOX2-GFP⁺ TECs. The few tumours that arise from the transplantation of a high number of SOX2-GFP⁻ TECs contained SOX2-GFP⁺ and SOX2-GFP⁻ cells (Extended Data Fig. 6g–j), suggesting that upon transplantation, rare SOX2-GFP⁻ cells can revert back to SOX2-GFP⁺ cells to sustain tumour propagation, but that this transition is very inefficient.

The higher proportion of TPCs in SOX2⁺ cells indicates that these cells may represent the cells that feed tumour growth within their natural environment. To investigate this possibility, we evaluated the impact of SOX2 cell lineage ablation on tumour growth (Extended Data Fig. 7a, b). Administering tamoxifen to SOX2CreER:Rosa-DTA mice with skin tumours led to the complete regression of most benign tumours (Fig. 2e, f and Extended Data Fig. 7c–d) and a strong regression of invasive SCCs (Fig. 2g and Extended Data Fig. 7e), suggesting that SOX2 marks a population of tumour cells necessary for tumour growth and maintenance *in vivo*.

To gain further insights into the molecular mechanisms that regulate CSC function, we transcriptionally profiled SOX2-GFP⁺ TECs. Interestingly, SOX2-GFP⁺ cells were enriched for genes known to be expressed by CSCs from other tumours or regulating stem cells in other systems (for example, *Hmga2* (refs 16, 17), *Ptprz1* (refs 18, 19), *Cd133* (also known as *Prom1*) (ref. 20), *Cd34* (refs 1–3), *Igf2bp2* (ref. 21) and *Vegfa* (ref. 3)) (Fig. 2h), proliferation (for example, *Ccnd1/2* and *Met*), cell survival and autophagy, cell adhesion (for example, *Itga3*), invasion, DNA damage response and resistance to therapy (for example, *Chek2*, *Aurka/b*, *Mgmt* and *Abcb1*), transcription factors (for example, *Pitx1* and *Etv5*) and chromatin regulators (for example, *Scmh1*) (Extended Data Fig. 8). In addition, a significant fraction of the genes upregulated

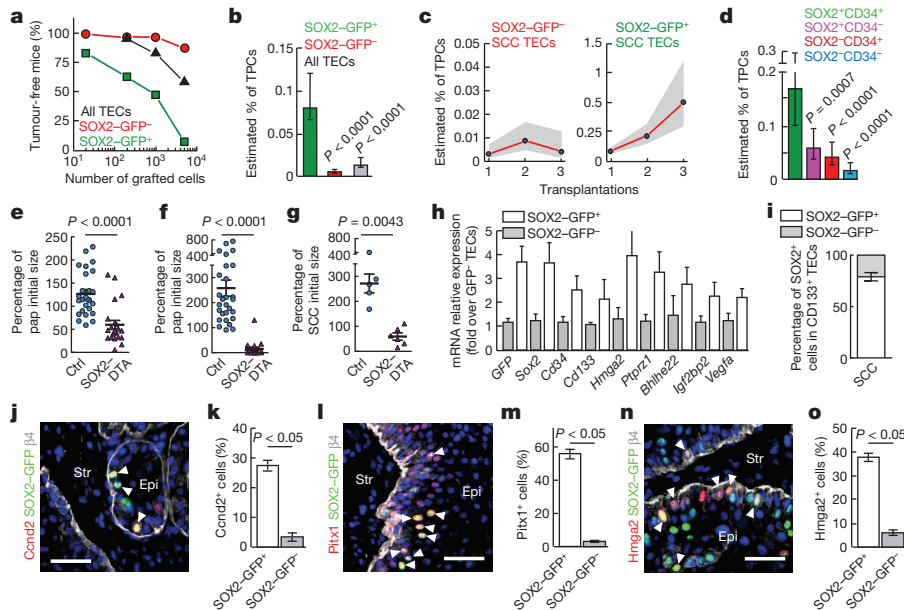


Figure 2 | SOX2 marks skin SCC tumour-propagating cells. **a**, Graph representing the percentage of tumour-free mice 5 weeks after subcutaneous injection of different dilutions of SOX2-GFP⁺, SOX2-GFP⁻ and all Epcam⁺ TECs into immunodeficient mice. **b**, Estimated percentage of TPCs in SOX2-GFP⁺ and SOX2-GFP⁻ TECs after the first transplantation using the extreme limiting dilution analysis³⁰ ($n = 30$ grafted tumours per dilution from five independent SCCs). **c**, Graph representing the estimated percentage of TPCs in SOX2-GFP⁺ and SOX2-GFP⁻ TECs during serial transplantations. **d**, Estimated percentage of TPCs in different TEC populations after the first transplantation ($n = 15$ grafted tumours per dilution from five independent SCCs). **e**, Quantification of the variation in tumour size 1 week after the beginning of tamoxifen administration (intraperitoneal (i.p.)) in control (Ctrl) and SOX2-DTA mice ($n > 22$ tumours from 4 different mice in each group).

in SOX2-GFP⁺ TECs were also preferentially expressed by embryonic epidermal cells (Extended Data Fig. 8), indicating that SOX2⁺ CSCs reactivate a genetic program expressed in the embryonic epidermis. Using FACS analysis and immunostaining, we confirmed the preferential expression of selected genes in SOX2-GFP⁺ CSCs, such as those controlling tumour proliferation (for example, *Ccnd2*) (Fig. 2j, k), *Pitx1*,

f, g Size of papilloma (pap; **f**) and SCC (**g**) 2 weeks after the beginning of tamoxifen administration (topical) in control and SOX2-DTA mice ($n > 18$ papillomas and $n > 5$ SCCs from 3 mice in each group). **h**, qRT-PCR analysis of stemness genes in SOX2-GFP⁺ and SOX2-GFP⁻ TECs ($n = 3$ tumours from 3 different mice for each group). **i**, FACS quantification of the percentage of SOX2-GFP⁺ cells in CD133⁺ TECs ($n = 5$ SCCs). **j–o**, Immunostaining and quantification of the expression of *Ccnd2* (**j, k**), *Pitx1* (**l, m**) and *Hmga2* (**n, o**) together with $\beta 4$ -integrin ($\beta 4$) in SOX2-GFP⁺ and SOX2-GFP⁻ TECs ($n \geq 2,578$ total cell counted per marker in $n = 5$ tumours from 3 different mice). Epi, epithelium; Str, tumour stroma. Hoechst nuclear staining is represented in blue; scale bars, 50 μ m. Data represent the mean and s.e.m. (except for transplantation assay: error bars (**b, d**) and grey area (**c**) represent the 95% confidence interval of the estimation).

a transcription factor expressed in human oral SCCs²² (Fig. 2l, m), cell adhesion (for example, *Itga3*) or associated with tumour stemness (for example, *Hmga2*, *Igf2bp2* and *Cd133*) (Fig. 2i, n, o and Extended Data Fig. 8).

SOX2 knockdown using short hairpin RNA (shRNA) in various cancer cell lines decreased tumour cell growth *in vitro* and/or *in vivo*

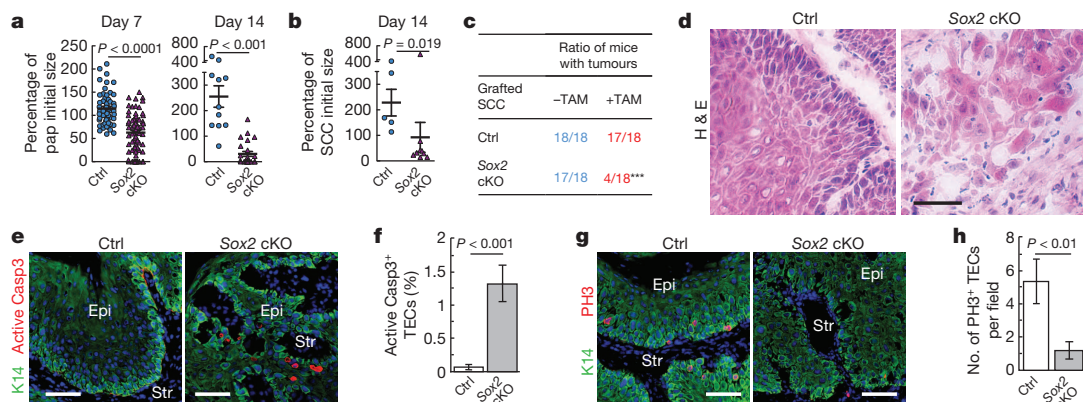
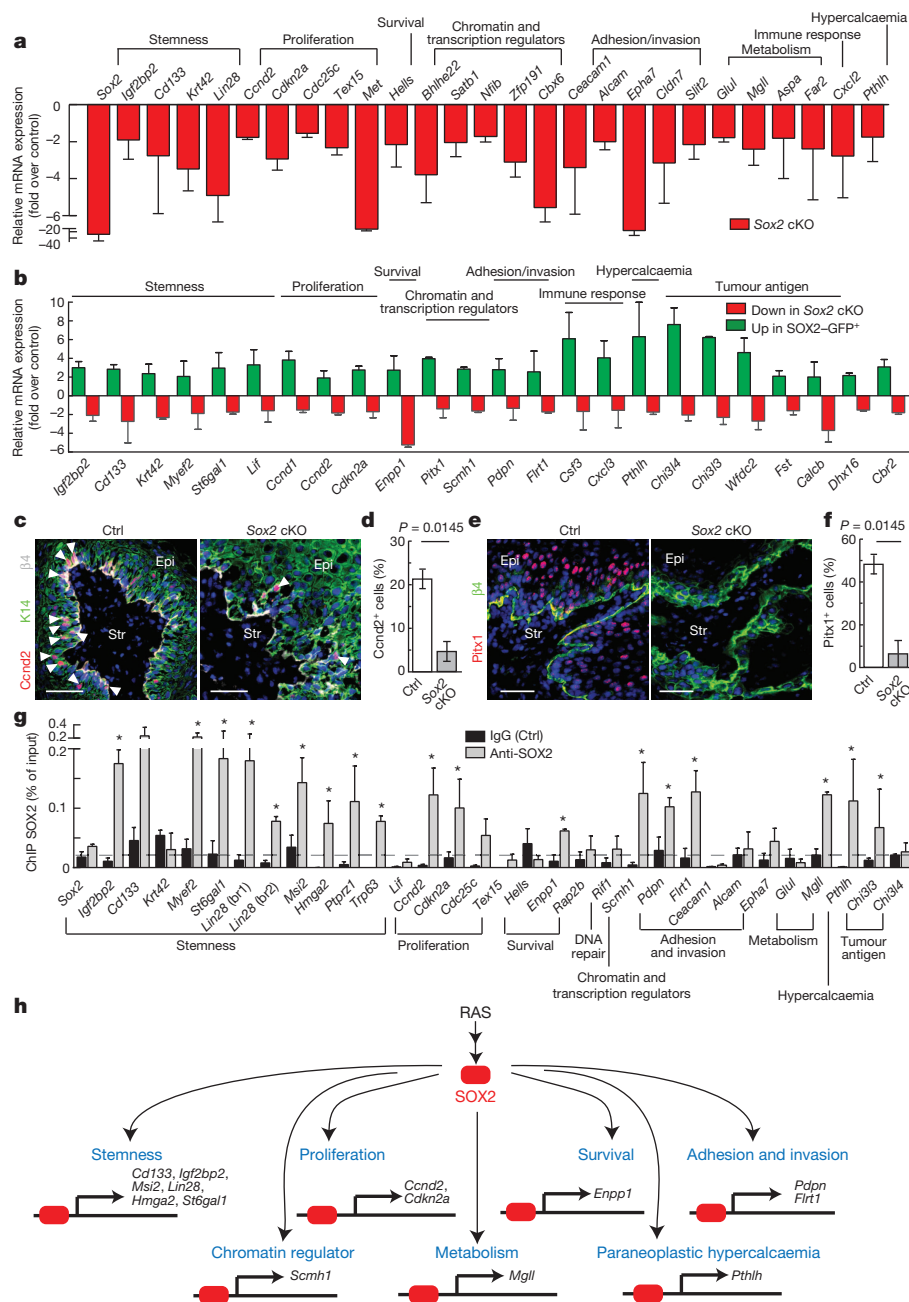


Figure 3 | SOX2 is essential for skin tumour maintenance. **a**, Tumour size 1 week (left) after the beginning of tamoxifen administration (12.5 mg i.p.) ($n \geq 6$ tumours from 8 different mice) and 2 weeks (right) after the beginning of topical tamoxifen administration (control $n = 11$ papillomas (pap); Sox2 conditional knockout $n = 24$ papillomas (pap) from 4 different mice in each group). **b**, Carcinoma size 2 weeks after the beginning of topical tamoxifen administration ($n > 5$ SCCs from 3 different mice). **c**, Table summarizing the frequency of secondary tumours after subcutaneous injection of 50,000 total SCC cells from control (Ctrl) and Sox2 conditional knockout (cKO) animals

after tamoxifen (TAM) administration ($n = 6$ grafted points per SCC). *** $P \leq 0.001$ (as compared with TAM-induced control). **d**, Haematoxylin and eosin (H&E) staining of control and Sox2 conditional knockout tumours. **e–h**, Immunostaining and quantification of TEC (K14⁺) apoptosis (active Casp3⁺) ($n = 7$ papillomas from 4 different mice) (**e, f**) and proliferation (phospho-histone H3⁺) ($n = 9$ papillomas from 3 different mice) (**g, h**) after Sox2 deletion. Epi, epithelium; Str, tumour stroma. Hoechst nuclear staining is represented in blue; scale bars, 50 μ m. Data represent the mean and s.e.m.

Figure 4 | SOX2 controls a gene network that regulates tumour proliferation and stemness.

a, qRT-PCR analysis of selected genes from different functional categories after Sox2 deletion ($n = 3$ tumours from 3 different mice for each genotype). **b**, Microarray analysis of genes upregulated (Up) in the SOX2⁺ CSC signature and downregulated (Down) in the SOX2-regulated gene signature, illustrating the genes positively controlled by SOX2 in CSCs ($n = 3$ different tumours from 3 mice in each group). Data represent the mean and s.e.m. of replicate arrays. **c–f**, Immunostaining and quantification of the expression of Ccnd2 (**c**, **d**) and Ptx1 (**e**, **f**) in K14 TECs after Sox2 deletion ($n \geq 1,743$ total cells counted in $n = 3$ papillomas from 3 different mice). Ctrl, control; cKO, conditional knockout; Epi, epithelium; Str, tumour stroma. Hoechst nuclear staining is represented in blue; scale bars, 50 μ m. **g**, SOX2 direct target genes in primary SCCs assessed by ChIP-qPCR analysis. Data represent the mean and s.e.m. of the percentage of input recovered after ChIP with SOX2 antibody or with the IgG at known²⁸ or putative SOX2-binding sites in the promoter region of the genes indicated. The dashed line represents the percentage of input for a negative control region (Trpm7) ($n = 3$ tumours from 3 different mice). br1/2, binding region1/2. * $P \leq 0.05$. **h**, Model of SOX2 functions in skin CSCs.



after their transplantation into immunodeficient mice²³. However, so far no study has demonstrated the requirement for SOX2 expression to sustain the growth of established primary tumours. To determine whether SOX2 regulates tumour maintenance and CSC functions *in vivo* in primary tumours, we performed conditional deletion of Sox2 in pre-existing skin tumours (Extended Data Fig. 9a, b). Tamoxifen administration over 2 weeks to K14CreER:SOX2^{fl/fl} mice with skin tumours led to complete regression of the vast majority of benign papillomas (Fig. 3a and Extended Data Fig. 9c) and induced a strong regression in the size of malignant SCCs (Fig. 3b). Consistent with the absence of SOX2 expression in a minor fraction of skin tumours, a few tumours did not regress after Sox2 deletion (Fig. 3a, b). These data show the essential role of SOX2 for primary tumour growth and maintenance *in vivo* in benign and malignant SCCs. Conditional deletion of Sox2 in SCC cells after their transplantation into immunodeficient mice decreased their TPC capacity by more than 80% (Fig. 3c), further supporting the notion that SOX2 regulates tumour stemness in primary skin SCCs. Sox2 deletion induced major changes in papilloma tumour-cell shape, which were much

bigger and flatter (Fig. 3d), markedly increased apoptosis and decreased cell proliferation in papilloma and SCC (Fig. 3e–h and Extended Data Fig. 10), leading to a decrease in the production of differentiated TECs (Extended Data Fig. 9e).

To determine the molecular mechanisms by which SOX2 regulates tumour maintenance and cancer-cell functions in primary mouse skin tumours, we induced Sox2 deletion in pre-existing tumour cells and performed microarray and qRT-PCR of TECs in the presence or absence of SOX2 (SOX2-regulated gene signature) (Fig. 4a). Functional analysis of these differentially expressed genes revealed that SOX2 regulates directly and/or indirectly the expression of a number of genes involved in many critical aspects of cancer functions²⁴ (Fig. 4a and Extended Data Fig. 9g).

To understand more precisely the mechanisms by which SOX2 regulates tumour stemness in SOX2-GFP⁺ CSCs, we defined the genes preferentially upregulated in SOX2-GFP⁺ cells and downregulated upon Sox2 deletion ($P = 6 \times 10^{-9}$) (Fig. 4b, Extended Data Fig. 9h and Supplementary Table 1). These genes included CSC markers (for example,

Cd133)²⁰, genes that control tumour stemness in other malignancies (for example, *Igf2bp2*)²¹, cell proliferation (for example, *Ccnd2*), survival (for example, *Enpp1*) and inflammation (for example, *Cxcl3*), transcription and chromatin remodelling (for example, *Pitx1* and *Scmh1*), as well as *Pthlh*, the gene responsible for paraneoplastic hypercalcaemia²⁵ (Fig. 4b). Immunostaining of papilloma in which *Sox2* had been acutely deleted confirmed that *Ccnd2*, a gene controlling cell proliferation, *Pitx1*, a transcription factor expressed in SCCs²², and *Pdpn*, a mucin-like glycoprotein expressed in a variety of SCCs associated with poor prognosis²⁶, were strongly decreased upon *Sox2* deletion (Fig. 4c–f and Extended Data Fig. 9f). Interestingly, a fraction of the downregulated genes following conditional deletion of *Sox2* in TECs are also found to be directly and indirectly regulated by SOX2 in embryonic stem cells^{27,28} and in brain tumour cell lines²⁹ (Extended Data Fig. 9h–j and Supplementary Tables 1–3), suggesting that SOX2 may regulate some common mechanisms across a range of stem cells including CSCs. Using ChIP-qPCR, we found that many genes downregulated upon *Sox2* deletion and controlling tumour proliferation and metabolism, and associated with stemness in other systems, such as *Ccnd2*, *Igf2bp2*, *Cd133*, *Scmh1*, *St6gal1*, *Mgl1*, *Pdpn*, *Msi2*, *Lin28* and *Trp63* were bound by SOX2 in primary SCC (Fig. 4g). Other SOX2-binding sites in *Glul*, *Hells* and *Chi3l4* (also known as *Chil4*), as well as in *Lif* and *Krt42*—which are directly bound by SOX2 in embryonic stem cells (Supplementary Tables 1, 2)—were not bound by SOX2 in primary SCCs (Fig. 4g), indicating the specificity of the SOX2-regulated genes in SCC. Our molecular analysis of SOX2 function in tumour cells shows that SOX2 directly regulates the expression of key genes involved in cancer-cell proliferation, stemness, chromatin regulation, metabolism and paraneoplastic hypercalcaemia in primary skin tumours *in vivo* (Fig. 4h).

We identify SOX2 as marking a continuum in skin carcinogenesis from tumour-initiating cells to the regulation of CSC functions in invasive cancer (Fig. 4h). Considering the broad diversity of cancers expressing SOX2, the functions and downstream target genes of SOX2 uncovered in primary skin tumours are likely to be relevant for other cancers, as well as for the development of novel strategies targeting CSCs.

METHODS SUMMARY

DMBA/TPA chemical carcinogenesis, measurement of tumour size, transplantation assays, immunostaining, RNA extraction and qRT-PCR were performed as previously described⁴. ChIP, Comparative genomic hybridization (CGH) array, FISH, microarray analysis and statistical analysis were performed as described in Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 August 2013; accepted 31 March 2014.

Published online 8 June 2014.

- Malanchi, I. *et al.* Cutaneous cancer stem cell maintenance is dependent on β -catenin signalling. *Nature* **452**, 650–653 (2008).
- Schober, M. & Fuchs, E. Tumor-initiating stem cells of squamous cell carcinomas and their control by TGF- β and integrin/focal adhesion kinase (FAK) signaling. *Proc. Natl Acad. Sci. USA* **108**, 10544–10549 (2011).
- Beck, B. *et al.* A vascular niche and a VEGF-Nrp1 loop regulate the initiation and stemness of skin tumours. *Nature* **478**, 399–403 (2011).
- Lapouge, G. *et al.* Skin squamous cell carcinoma propagating cells increase with tumour progression and invasiveness. *EMBO J.* **31**, 4563–4575 (2012).
- Sarkar, A. & Hochedlinger, K. The Sox family of transcription factors: versatile regulators of stem and progenitor cell fate. *Cell Stem Cell* **12**, 15–30 (2013).
- Arnold, K. *et al.* Sox2⁺ adult stem and progenitor cells are important for tissue regeneration and survival of mice. *Cell Stem Cell* **9**, 317–329 (2011).
- Bass, A. J. *et al.* SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nature Genet.* **41**, 1238–1242 (2009).
- Alam, M. & Ratner, D. Cutaneous squamous-cell carcinoma. *N. Engl. J. Med.* **344**, 975–983 (2001).
- Ellis, P. *et al.* SOX2, a persistent marker for multipotential neural stem cells derived from embryonic stem cells, the embryo or the adult. *Dev. Neurosci.* **26**, 148–165 (2004).

- Laga, A. C. *et al.* Expression of the embryonic stem cell transcription factor SOX2 in human skin: relevance to melanocyte and Merkel cell biology. *Am. J. Pathol.* **176**, 903–913 (2010).
- Driskell, R. R., Giangreco, A., Jensen, K. B., Mulder, K. W. & Watt, F. M. Sox2-positive dermal papilla cells specify hair follicle type in mammalian epidermis. *Development* **136**, 2815–2823 (2009).
- Bardot, E. S. *et al.* Polycomb subunits Ezh1 and Ezh2 regulate the Merkel cell differentiation program in skin stem cells. *EMBO J.* **32**, 1990–2000 (2013).
- Taranova, O. V. *et al.* SOX2 is a dose-dependent regulator of retinal neural progenitor competence. *Genes Dev.* **20**, 1187–1202 (2006).
- Beck, B. & Blanpain, C. Unravelling cancer stem cell potential. *Nature Rev. Cancer* **13**, 727–738 (2013).
- Meacham, C. E. & Morrison, S. J. Tumour heterogeneity and cancer cell plasticity. *Nature* **501**, 328–337 (2013).
- Copley, M. R. *et al.* The Lin28b–let-7–Hmga2 axis determines the higher self-renewal potential of fetal haematopoietic stem cells. *Nature Cell Biol.* **15**, 916–925 (2013).
- Nishino, J., Kim, I., Chada, K. & Morrison, S. J. Hmga2 promotes neural stem cell self-renewal in young but not old mice by reducing p16^{Ink4a} and p19^{Arf} expression. *Cell* **135**, 227–239 (2008).
- Himburg, H. A. *et al.* Pleiotrophin regulates the retention and self-renewal of hematopoietic stem cells in the bone marrow vascular niche. *Cell Rep.* **2**, 964–975 (2012).
- Soh, B. S. *et al.* Pleiotrophin enhances clonal growth and long-term expansion of human embryonic stem cells. *Stem Cells* **25**, 3029–3037 (2007).
- Grosse-Gehling, P. *et al.* CD133 as a biomarker for putative cancer stem cells in solid tumours: limitations, problems and challenges. *J. Pathol.* **229**, 355–378 (2013).
- Janiszewska, M. *et al.* Imp2 controls oxidative phosphorylation and is crucial for preserving glioblastoma cancer stem cells. *Genes Dev.* **26**, 1926–1944 (2012).
- Libório, T. N. *et al.* *In situ* hybridization detection of homeobox genes reveals distinct expression patterns in oral squamous cell carcinomas. *Histopathology* **58**, 225–233 (2011).
- Liu, K. *et al.* The multiple roles for Sox2 in stem cell maintenance and tumorigenesis. *Cell. Signal.* **25**, 1264–1271 (2013).
- Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
- Esbrit, P. Hypercalcaemia of malignancy—new insights into an old syndrome. *Clin. Lab.* **47**, 67–71 (2001).
- Chuang, W. Y., Chang, Y. S., Yeh, C. J., Wu, Y. C. & Hsueh, C. Role of podoplanin expression in squamous cell carcinoma of upper aerodigestive tract. *Histol. Histopathol.* **28**, 293–299 (2013).
- Masui, S. *et al.* Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature Cell Biol.* **9**, 625–635 (2007).
- Marson, A. *et al.* Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**, 521–533 (2008).
- Fang, X. *et al.* The SOX2 response program in glioblastoma multiforme: an integrated ChIP-seq, expression microarray, and microRNA analysis. *BMC Genomics* **12**, 11 (2011).
- Hu, Y. & Smyth, G. K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J. Immunol. Methods* **347**, 70–78 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank our colleagues who provided us with reagents. We also thank the animal house facility of the Université Libre de Bruxelles (ULB) (Erasmus campus). C.B. is an investigator of WELBIO. S.Bo., B.B., B.D., S.Br., G.L. and D.N. are supported by a FNRS/FRIA, FNRS and TELEVIE fellowships. G.D. is supported by the Brussels Region through the BB2B program. This work was supported by the FNRS, TELEVIE, BB2B program, the IUAP program, a research grant from the Fondation Contre le Cancer, the ULB foundation, the Fonds Yvonne Boël, the Fonds Gaston Ithier, the foundation Bettencourt Schueller, and a starting grant from the European Research Council.

Author Contributions C.B., S.Bo., G.D. and G.L. designed the experiments and performed data analysis. S.Bo., G.L., G.D., D.N. and B.B. performed all the experiments. V.d.M., S.R., M.L.M. and I.S. collected data and performed the analysis of SOX2 expression and gene amplification on human samples. A.C., E.N., S.L. and C.D. provided technical support. S.Br. performed microarray analysis. B.D. and F.F. performed and analysed ChIP for histone marks. C.B. wrote the manuscript.

Author Information Microarray data have been deposited in the Gene Expression Omnibus under accession numbers GSE55737 and GSE55738. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.B. (Cedric.Blanpain@ulb.ac.be).

Avoidance of ribonucleotide-induced mutations by RNase H2 and Srs2–Exo1 mechanisms

Catherine J. Potenski¹*, Hengyao Niu²*, Patrick Sung² & Hannah L. Klein¹

Srs2 helicase is known to dismantle nucleofilaments of Rad51 recombinase to prevent spurious recombination events^{1–6} and unwind trinucleotide sequences that are prone to hairpin formation⁷. Here we document a new, unexpected genome maintenance role of Srs2 in the suppression of mutations arising from mis-insertion of ribonucleoside monophosphates during DNA replication. In cells lacking RNase H2, Srs2 unwinds DNA from the 5' side of a nick generated by DNA topoisomerase I⁸ at a ribonucleoside monophosphate residue. In addition, Srs2 interacts with and enhances the activity of the nuclease Exo1, to generate a DNA gap in preparation for repair. Srs2–Exo1 thus functions in a new pathway of nick processing-gap filling that mediates tolerance of ribonucleoside monophosphates in the genome. Our results have implications for understanding the basis of Aicardi–Goutières syndrome, which stems from inactivation of the human RNase H2 complex⁹.

Mutants of *SRS2* (*hpr5*) and *RNH202* (*hpr4*), which encode a 3' to 5' helicase and a subunit of the RNase H2 complex, respectively, are represented in our collection of hyper-recombination mutants¹⁰. These mutants show similar hyper-recombination levels¹⁰, DNA damage sensitivity, and interactions with DNA replication and damage repair genes (Extended Data Fig. 1). The *rnh202Δ srs2Δ* double mutant was tested to determine whether RNase H2 and Srs2 might act together. A truncation mutant, *srs2-860* (ref. 1), that has helicase activity but lacks Rad51 interaction was used (Fig. 1a), as we wished to distinguish between DNA helicase and Rad51 removal activities *in vivo*. We disabled the ATP hydrolysis activity of *srs2-860* through the K41A mutation (*srs2-KA860*)¹¹. The *rnh202Δ srs2-KA860* cells grew slowly (Fig. 1b), showing that Srs2 helicase activity is important for growth in *rnh202Δ*. Similar results were obtained with the *rnh201Δ srs2Δ* mutant, deficient in the catalytic subunit of RNase H2 (data not shown). We note that the *srs2-KA* mutation in the full-length protein¹¹ also engenders a slow growth phenotype with *rnh202Δ* (Extended Data Fig. 2a, b).

Cells deficient in *RNH202*, but not *SRS2*, accumulate spontaneous mutations¹⁰. Interestingly, the *rnh202 srs2* double mutant shows a much stronger mutator phenotype (Fig. 1c), indicating that *SRS2* prevents mutations in *rnh202* cells. The *srs2-860* allele possesses this mutation avoidance function, but *srs2-KA860* is defective in this regard. The *srs2-860* mutant has a hyper-recombination phenotype because its encoded protein lacks anti-Rad51 activity¹, and the *rnh202 srs2-860* double mutant shows an additive increase in recombination that is partly dependent on the Srs2 helicase activity (Fig. 1d). Consistent with this finding, the *rnh202 srs2* mutant shows increased Rad52 foci (Extended Data Fig. 3a), indicative of spontaneous DNA double strand breaks¹². The slow growth phenotype of the *rnh202 srs2* mutant (Fig. 1b) is due to an increase in G2 cells and can be partly relieved by inactivating the DNA damage checkpoints (Extended Data Fig. 3b, c). Lastly, the DNA damage checkpoints are largely intact in these cells, as shown by Rad53 phosphorylation after hydroxyurea treatment (Extended Data Fig. 3d), which elevates the rNTP:dNTP ratio. However, combining *rnh202Δ* with either *srs2Δ* or *srs2-KA860* leads to a synergistic sensitivity to hydroxyurea (Fig. 1e).

Ribonucleotide misincorporation becomes elevated in cells that harbour the *pol2M644G* mutation in the replicative DNA polymerase Polε¹³, and the combination of *rnh202* with *pol2M644G* increases the ribonucleotide load further¹⁴. Importantly, the *rnh202 srs2 pol2M644G* triple mutation is lethal (Fig. 1f), and the lethality is rescued by *srs2-860* but not *srs2-KA860* (data not shown).

Loss of RNase H2 causes slippage mutations in simple repeats without affecting base substitutions^{8,14}. We sequenced *can1* alleles to determine the mutation signatures of various strains. As expected, *rnh202Δ* cells accumulate slippage mutations, including single nucleotide deletions in a mononucleotide run and dinucleotide deletions in dinucleotide repeats (Fig. 1g and Extended Data Table 1a). The *srs2Δ* mutant

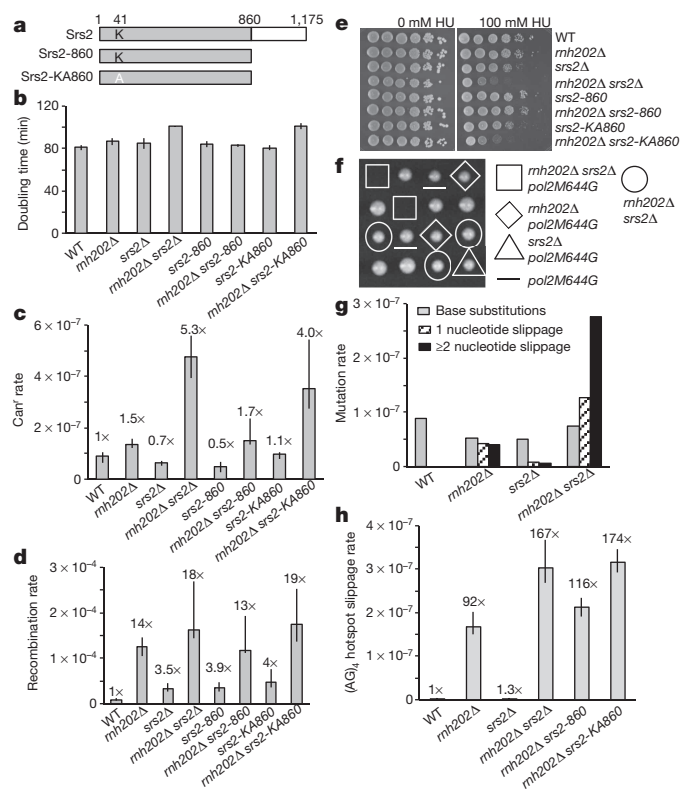


Figure 1 | Genome stability maintenance role of Srs2 helicase in *rnh202Δ* cells. **a**, Strains harbouring either the *srs2-860* allele or its helicase null K41A variant at the chromosomal *SRS2* locus were constructed. **b**, Doubling times. WT, wild type. **c**, Spontaneous mutation rates. **d**, Recombination rates. **e**, Sensitivity to hydroxyurea (HU). **f**, Synthetic interaction of the *pol2M644G* mutation in the *rnh202Δ srs2Δ* background. **g**, Mutation spectra of *Can1* mutants obtained in **c**. **h**, Dinucleotide slippage rates. Plotted in **b** are the mean values \pm s.d. ($n = 3$). Plotted in **c**, **d** and **h** are median rates with error bars representing 95% confidence intervals²³ ($n = 18$).

¹Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, New York 10016, USA. ²Molecular Biophysics and Biochemistry, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06520, USA.

*These authors contributed equally to this work.

shows a slight increase in these mutation types, consistent with the role of Srs2 in preventing slippage in long trinucleotide repeats⁷. The slippage mutations are increased in the *rnh202 srs2* double mutant (Fig. 1g and Extended Data Table 1a). To characterize this mutator phenotype further, we used a reporter that detects slippages in AG repeats⁸. With this reporter, *srs2Δ* is indistinguishable from wild type, whereas *rnh202Δ* shows a 90-fold increase. This phenotype of *rnh202Δ* is exacerbated by removing Srs2 or inactivating its helicase activity (Fig. 1h and Extended Data Fig. 2c). Altogether, the results show that Srs2 helicase guards against mutations induced by ribonucleoside monophosphates (rNMPs) in DNA, and they show an overlap in mutation sequences between *rnh202* and *rnh202 srs2* with new slippage sites of at least three nucleotides in the double mutant (Extended Data Fig. 4 and Extended Data Table 1a).

Mutations that arise in RNase-H2-deficient cells are dependent on topoisomerase I (Topo I), encoded by the *TOP1* gene, which cleaves DNA at rNMP residues^{8,15}. We tested *top1Δ* to ask whether it is relevant for *rnh202 srs2*. Can^r mutations arise from different events, including gross chromosome rearrangements¹⁶. In our test, *top1Δ* is epistatic to *rnh202* (Fig. 2a). Importantly, *rnh202 top1* shows a specific loss of slippage events of two nucleotides or more (Fig. 2b and Extended Data Table 1a). The increased mutation rate in *rnh202 srs2* is also suppressed by *top1Δ* (Fig. 2a), including a complete loss of slippage events of at least two nucleotides and a reduction in single nucleotide slippage (Fig. 2b and Extended Data Table 1a). The hydroxyurea hypersensitivity (Fig. 2c), the hyper-recombination phenotype (Fig. 2d) and the lethality with *pol2M644G* (Fig. 2e) of *rnh202 srs2* are also overcome by *top1Δ*. Thus, when topoisomerase I is present, the tolerance of rNMPs needs RNase H2 and Srs2, and recombinogenic DNA double strand breaks arise when RNase H2 is absent.

Topo-I-mediated cleavage of an rNMP results in unligatable 5' hydroxyl and 3' cyclic phosphate termini¹⁵. We reasoned that, in mutation avoidance, Srs2 unwinds DNA from the nick generated by Topo I. To test this, we constructed linear duplex DNA substrates whose ends were occluded by a streptavidin–biotin complex. As expected, the intact substrate was resistant to unwinding by Srs2 (Extended Data Fig. 5). Importantly, when a nick bearing a 5' OH group was introduced, ATP-dependent displacement of the strand containing this group occurred (Extended

Data Fig. 5). In contrast, little unwinding of the 3' OH-bearing strand was seen (Extended Data Fig. 5).

To examine whether Srs2 unwinds DNA from a nick introduced by Topo I at an embedded ribonucleotide, we prepared a DNA substrate with a single uridine monophosphate (UMP) residue and blocked ends. As expected¹⁵, calf thymus Topo I, but not *Escherichia coli* Topo I, cleaved the DNA/ribonucleoside UMP (rUMP) hybrid strand, but not an equivalent substrate without the ribonucleotide (Extended Data Fig. 6a, b). Incubation of the Topo-I-nicked DNA (Extended Data Fig. 6c) with Srs2 led to the ATP-dependent displacement of the strand with the 5' OH group, but not that containing the 3' cyclic phosphate (Fig. 3a, b). Mutations in *MPH1*, which encodes a 3'–5' helicase, also lead to a mutator phenotype¹⁷. However, Mph1 protein does not act on the Topo-I-nicked substrate (Extended Data Fig. 7a). Consistent with this, the *rnh202Δ mph1Δ* double mutant shows the same dinucleotide slippage rate as *rnh202Δ* (Extended Data Fig. 7b).

We surmised that Srs2-mediated unwinding of the 5'-strand at the Topo-I-induced DNA nick prevents re-ligation that is associated with DNA deletion and serves to initiate lesion removal by creating a DNA gap (see discussions later). If this were the case, one would expect a 5' endo- or exonuclease to be involved in the removal of the displaced DNA strand. One strong candidate is Exo1, which has 5' to 3' exonuclease and 5' Flap endonuclease activities and plays an important role in DNA repair, including DNA break end resection during homologous recombination and DNA digestion during mismatch repair. Importantly, with the dinucleotide slippage reporter, we found that Exo1 has an antitumor activity against *rnh202*, and that it acts in an epistatic manner with Srs2 (Fig. 3c).

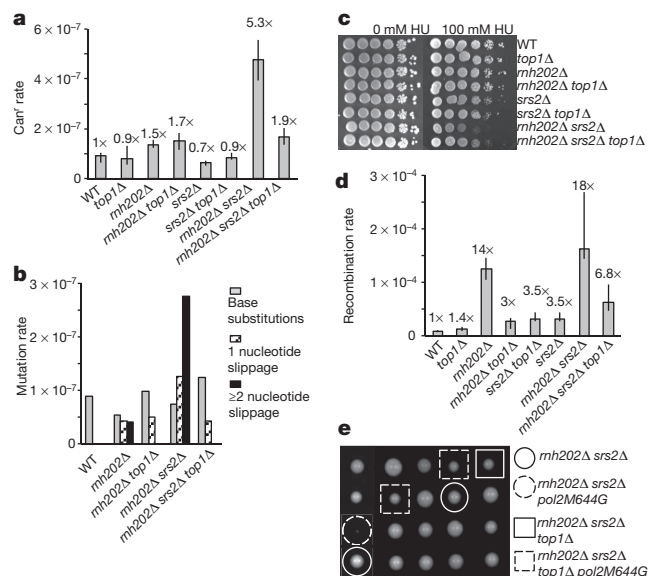


Figure 2 | Causation of genome instability in *rnh202Δ srs2Δ* mutant by Top1. **a**, Spontaneous mutation rates. **b**, Mutation spectra of Can^r mutants obtained in **a**. **c**, Sensitivity to hydroxyurea. **d**, Recombination rates. **e**, Synthetic interaction of the *pol2M644G* mutation in the *rnh202Δ srs2Δ top1Δ* background. Plotted in **a** and **d** are median rates with error bars representing 95% confidence intervals²³ ($n = 18$).

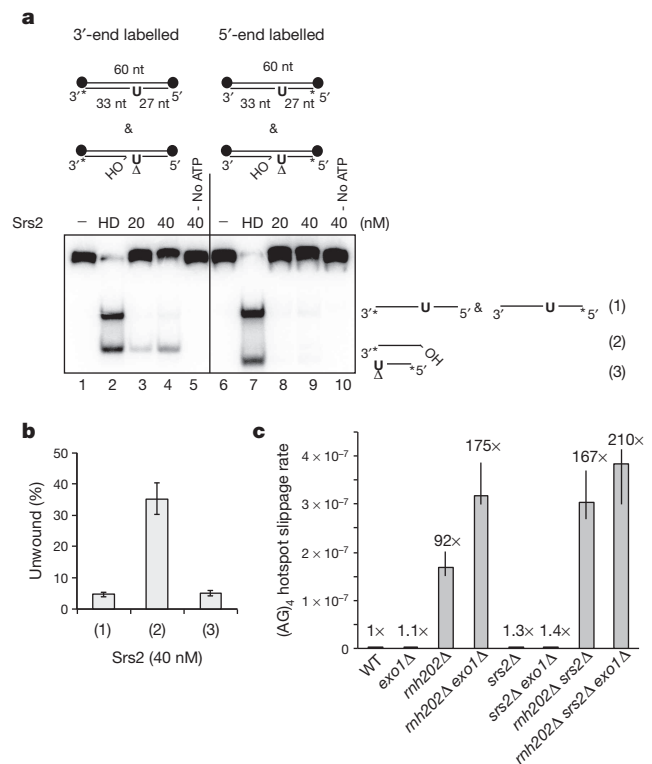


Figure 3 | Srs2-mediated displacement of the 5' strand at a Topo-I-induced DNA nick. **a**, **b**, DNA substrates containing a Topo-I-induced nick and labelled at either the 3' or 5' end were incubated with Srs2 with or without ATP. The reactions were analysed on a polyacrylamide gel (a); the results were quantified (b). Heat-denatured (HD) DNA was analysed in lanes 2 and 7. nt, nucleotide. **c**, Effects of Exo1 deletion on slippage mutations in *rnh202Δ* and *rnh202Δ srs2Δ* mutant cells. Dinucleotide slippage rates. Plotted in **b** are the mean values \pm s.d. ($n = 3$); in **c**, median rates with error bars representing 95% confidence intervals²³ ($n = 18$).

Being prompted by the genetic results, we next tested possible functional interactions between Srs2 and Exo1. For this, DNA substrates blocked at their ends and harbouring a 5' Flap or a 5' Flap adjoining a gap were constructed; the latter substrate resembles the intermediate generated by Srs2-mediated DNA unwinding from a nick. We found that Exo1 acts endonucleolytically on these substrates, incising the single-stranded arm of the Flap structure (Extended Data Fig. 7c); little exonuclease activity was detected with these substrates or with an oligonucleotide dT substrate (Extended Data Fig. 7c). Importantly, Srs2 enhanced the endonucleolytic cleavage of the gapped 5' Flap DNA in an ATP-independent manner (Fig. 4a, b). Interestingly, by affinity pull-down, a physical interaction between Srs2 and Exo1 was detected (Fig. 4c). In contrast, we did not detect any enhancement of Exo1 activity by Mph1 (Extended Data Fig. 8a). Moreover, although human EXO1 (hEXO1) digests the 5' Flap substrate similarly to yeast Exo1, addition of Srs2 led to inhibition of the reaction (Extended Data Fig. 8b). We also tested the effect of Srs2 on 5' Flap cleavage mediated by yeast Dna2 and human FEN1 and found that Srs2 attenuates the cleavage efficiency in both cases (data not shown). These results support the premise that Srs2 and Exo1 cooperate, in a specific fashion, in the downstream processing of the DNA nick originating from Topo-I-mediated cleavage at an rNMP.

hEXO1 can digest DNA exonucleolytically with 5'–3' polarity at a nick¹⁸, and we verified that yeast Exo1 has the same activity (Extended Data Fig. 7c). To ask how Exo1 processes different DNA nick types, we prepared substrates that harbour a Topo-I-induced nick, a nick with 3' hydroxyl and 5' phosphate groups, or a nick with 3' and 5' hydroxyl groups. As shown in Fig. 4d–g, Exo1 can process the Topo-I-induced nick, although less efficiently than it does a clean DNA nick. Interestingly, Srs2 enhanced digestion of the Topo-I-induced nick, by as much as fourfold (Fig. 4d, g). A stimulatory effect of Srs2, although less pronounced, on Exo1 was also observed with the nicked substrate harbouring a 5'-OH group (Fig. 4e, g), but little enhancement occurred at the nick harbouring the 3' OH and 5' phosphate groups (Fig. 4f, g). Again,

ATP is dispensable for the observed stimulation (Fig. 4d–g), which suggests Srs2 facilitates the delivery of Exo1 to the cleaved rNMP site. We found that Mph1 is unable to stimulate Exo1-catalysed digestion of the Topo-I-induced nick (Extended Data Fig. 8c), and that digestion by hEXO1 is attenuated by Srs2 (Extended Data Fig. 8d).

The genetic and biochemical data presented above support the model (Extended Data Fig. 9) where Srs2 unwinds DNA from the Topo-I-generated nick and recruits Exo1, which uses its 5' exonuclease or 5' Flap endonuclease activity to generate a DNA gap to be filled by a DNA polymerase. In the absence of Srs2, unwinding of DNA from the 5' OH end is not possible and its processing by Exo1 becomes inefficient. In this case, Topo I could act again at the 3' cyclic phosphate, cleaving two nucleotides and re-ligating the ends to generate a two-nucleotide deletion¹⁹.

Recent evidence has suggested a beneficial role of DNA nicks stemming from ribonucleotide excision repair in the facilitation of DNA mismatch removal from the newly replicated DNA^{20,21}. However, rNMPs in DNA can be harmful, as they cause slippage and other mutations. To avoid such mutations, RNase H2 initiates rNMP elimination by ribonucleotide excision repair²². We have shown that in the absence of RNase H2, the main mechanism to remove embedded rNMPs is cleavage by Topo I and nick processing by Srs2–Exo1, followed by gap repair. rNMPs cleaved by Topo I may also be converted into a recombinogenic double strand break. Our work thus uncovers a new function of Srs2 helicase and Exo1 in mutation avoidance and rNMP lesion repair. Because the Srs2 helicase activity is required for the avoidance of rNMP-induced mutations, the main repair mechanism in cells may entail Srs2-mediated DNA strand separation from the Topo-I-induced nick, to create a 5' Flap for Exo1 to incise. The physical interaction between Srs2 and Exo1 probably serves to recruit the latter to the Flap structure.

METHODS SUMMARY

Yeast strains are listed in Extended Data Table 1b.

Mutation rates were determined by the Lea and Coulson fluctuation method. Independent mutants were sequenced to identify mutations. Recombination assays were performed using the *leu2-ecoRI::URA3::leu2-bstEII* system as described¹⁰. For the (AG)₄ reporter, rates were determined as described^{18,23}. At least 18 independent cultures (minimum of two isolates per genotype) were used to determine rates, and 95% confidence intervals were determined as described²³.

In vitro assays. Srs2 and Exo1 were expressed in *E. coli* and Hi5 insect cells, respectively, and purified as described^{24,25}. Purified hEXO1 was a gift from G.-M. Li. Nuclease and helicase assays were conducted as described elsewhere (see Methods). Reactions were analysed by gel electrophoresis and phosphorimaging.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 November 2013; accepted 28 March 2014.

Published online 1 June 2014.

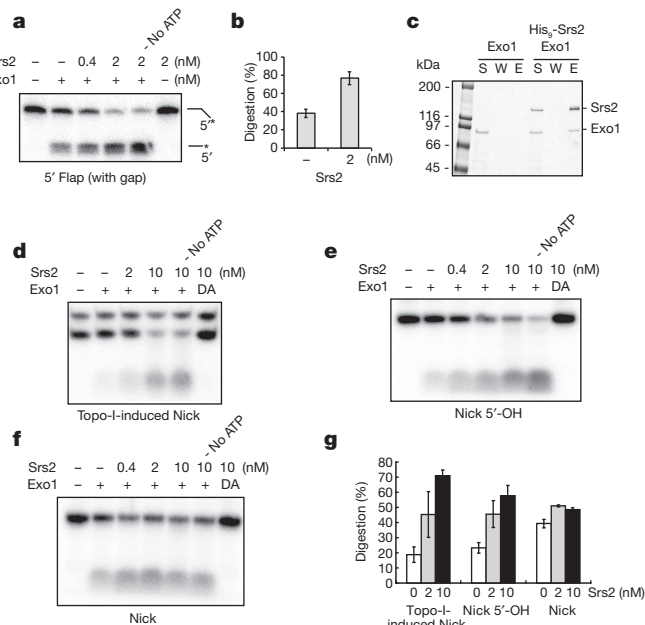


Figure 4 | Interactions of Exo1 with Srs2. **a, b**, Endonucleolytic cleavage of a gapped 5' Flap DNA by Exo1 (0.1 nM) was examined with and without Srs2 or ATP. The reactions were analysed on a polyacrylamide gel (**a**) and the results were quantified (**b**). **c**, Affinity pull-down through the (His)₆ tag on Srs2 was used to analyse its interaction with Exo1. The supernatant (S), wash (W) and SDS eluate (E) fractions were analysed by 4–20% SDS–polyacrylamide gel electrophoresis. **d–g**, Exonucleolytic cleavage of a (**d**) Topo-I-induced nick, (**e**) nick bearing a 5'-OH group and (**f**) clean nick bearing a 5'-phosphate by Exo1 was examined with and without Srs2 or ATP. The reactions were analysed as above (**g**). Plotted in **b** and **g** are the mean values \pm s.d. ($n = 3$).

- Colavito, S. *et al.* Functional significance of the Rad51–Srs2 complex in Rad51 presynaptic filament disruption. *Nucleic Acids Res.* **37**, 6754–6764 (2009).
- Krejci, L. *et al.* DNA helicase Srs2 disrupts the Rad51 presynaptic filament. *Nature* **423**, 305–309 (2003).
- Van Komen, S., Reddy, M. S., Krejci, L., Klein, H. & Sung, P. ATPase and DNA helicase activities of the *Saccharomyces cerevisiae* anti-recombinase Srs2. *J. Biol. Chem.* **278**, 44331–44337 (2003).
- Aboussekhra, A., Chanet, R., Adjiri, A. & Fabre, F. Semidominant suppressors of Srs2 helicase mutations of *Saccharomyces cerevisiae* map in the RAD51 gene, whose sequence predicts a protein with similarities to procaryotic RecA proteins. *Mol. Cell. Biol.* **12**, 3224–3234 (1992).
- Chanet, R., Heude, M., Adjiri, A., Maloisel, L. & Fabre, F. Semidominant mutations in the yeast Rad51 protein and their relationships with the Srs2 helicase. *Mol. Cell. Biol.* **16**, 4782–4789 (1996).
- Veaute, X. *et al.* The Srs2 helicase prevents recombination by disrupting Rad51 nucleoprotein filaments. *Nature* **423**, 309–312 (2003).
- Bhattacharyya, S. & Lahue, R. S. *Saccharomyces cerevisiae* Srs2 DNA helicase selectively blocks expansions of trinucleotide repeats. *Mol. Cell. Biol.* **24**, 7324–7330 (2004).
- Kim, N. *et al.* Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase I. *Science* **332**, 1561–1564 (2011).
- Crow, Y. J. *et al.* Mutations in genes encoding ribonuclease H2 subunits cause Aicardi-Goutières syndrome and mimic congenital viral brain infection. *Nature Genet.* **38**, 910–916 (2006).

10. Aguilera, A. & Klein, H. L. Genetic control of intrachromosomal recombination in *Saccharomyces cerevisiae*. I. Isolation and genetic characterization of hyper-recombination mutations. *Genetics* **119**, 779–790 (1988).
11. Krejci, L. *et al.* Role of ATP hydrolysis in the antirecombinase function of *Saccharomyces cerevisiae* Srs2 protein. *J. Biol. Chem.* **279**, 23193–23199 (2004).
12. Lisby, M., Rothstein, R. & Mortensen, U. H. Rad52 forms DNA repair and recombination centers during S phase. *Proc. Natl Acad. Sci. USA* **98**, 8276–8282 (2001).
13. Nick McElhinny, S. A. *et al.* Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proc. Natl Acad. Sci. USA* **107**, 4949–4954 (2010).
14. Nick McElhinny, S. A. *et al.* Genome instability due to ribonucleotide incorporation into DNA. *Nature Chem. Biol.* **6**, 774–781 (2010).
15. Sekiguchi, J. & Shuman, S. Site-specific ribonuclease activity of eukaryotic DNA topoisomerase I. *Mol. Cell* **1**, 89–97 (1997).
16. Chen, C., Umezu, K. & Kolodner, R. D. Chromosomal rearrangements occur in *S. cerevisiae* rfa1 mutator mutants due to mutagenic lesions processed by double-strand-break repair. *Mol. Cell* **2**, 9–22 (1998).
17. Scheller, J., Schurer, A., Rudolph, C., Hettwer, S. & Kramer, W. MPH1, a yeast gene encoding a DEAH protein, plays a role in protection of the genome from spontaneous and chemically induced damage. *Genetics* **155**, 1069–1081 (2000).
18. Genschel, J., Bazemore, L. R. & Modrich, P. Human exonuclease I is required for 5' and 3' mismatch repair. *J. Biol. Chem.* **277**, 13302–13311 (2002).
19. Cho, J. E., Kim, N., Li, Y. C. & Jinks-Robertson, S. Two distinct mechanisms of Topoisomerase I-dependent mutagenesis in yeast. *DNA Repair* **12**, 205–211 (2013).
20. Ghodgaonkar, M. M. *et al.* Ribonucleotides misincorporated into DNA act as strand-discrimination signals in eukaryotic mismatch repair. *Mol. Cell* **50**, 323–332 (2013).
21. Lujan, S. A., Williams, J. S., Clausen, A. R., Clark, A. B. & Kunkel, T. A. Ribonucleotides are signals for mismatch repair of leading-strand replication errors. *Mol. Cell* **50**, 437–443 (2013).
22. Sparks, J. L. *et al.* RNase H2-initiated ribonucleotide excision repair. *Mol. Cell* **47**, 980–986 (2012).
23. Spell, R. M. & Jinks-Robertson, S. Determination of mitotic recombination rates by fluctuation analysis in *Saccharomyces cerevisiae*. *Methods Mol. Biol.* **262**, 3–12 (2004).
24. Adkins, N. L., Niu, H., Sung, P. & Peterson, C. L. Nucleosome dynamics regulates DNA processing. *Nature Struct. Mol. Biol.* **20**, 836–842 (2013).
25. Seong, C., Colavito, S., Kwon, Y., Sung, P. & Krejci, L. Regulation of Rad51 recombinase presynaptic filament assembly via interactions with the Rad52 mediator and the Srs2 anti-recombinase. *J. Biol. Chem.* **284**, 24363–24371 (2009).

Acknowledgements We thank S. Jinks-Robertson and T. Kunkel for plasmids used in strain construction, R. Bermejo and M. Foiani for the genome sequencing of the original *hpr4* strain, and G.-M. Li and S. Lee for providing hEXO1. This work was supported by National Institutes of Health grants RO1GM053738 (to H.L.K.), RO1ES007061 (to P.S.) and K99ES021441 (to H.N.).

Author Contributions H.L.K. and P.S. conceived the study. C.J.P., H.N., P.S. and H.L.K. designed the experiments and analysed the data. C.J.P. and H.N. executed the experiments. C.J.P., H.N., P.S. and H.L.K. wrote the paper.

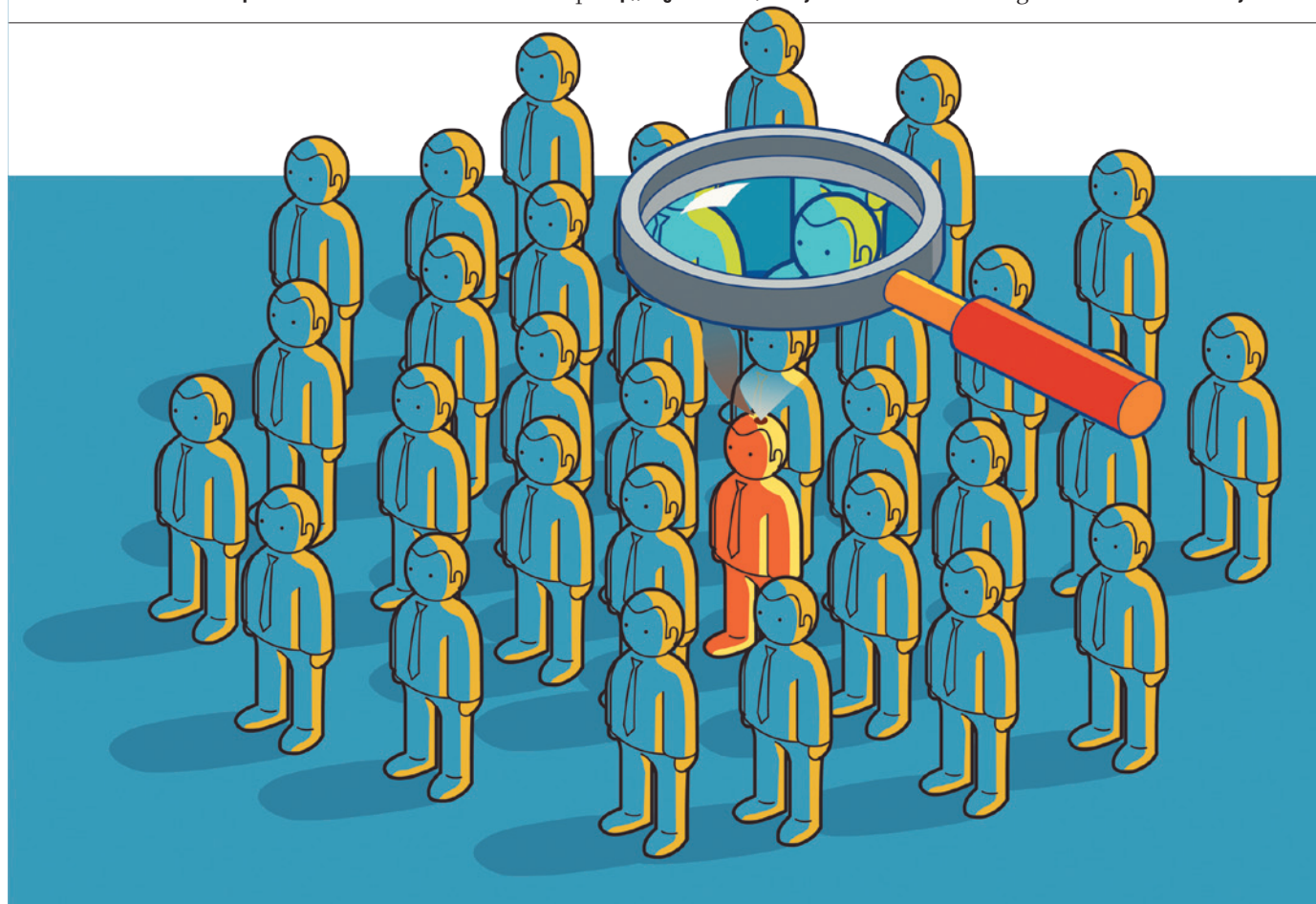
Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.S. (patrick.sung@yale.edu) or H.K. (Hannah.Klein@nyumc.org).

CAREERS

COLUMN Academia is the new 'alternative' for life-science careers **p.256**

NATUREJOBS BLOG The latest science-careers news and tips <http://blogs.nature.com/naturejobs>

NATUREJOBS For the latest career listings and advice www.naturejobs.com



EX15/ISTOCK VECTORS/GETTY

EMPLOYMENT

PhD overdrive

An excess of graduates means that job-seekers need to be versatile.

BY PAUL SMAGLIK

The low unemployment rate for people with a PhD in science, technology, engineering or mathematics in the United States is misleading. An increase in the number of such scientists — from both within the country and elsewhere — over the past decade seems to have been absorbed by the US workforce. But observers warn that the demand is reaching saturation point, especially in the biomedical-research sector.

The number of PhD scientists and engineers either in employment or actively seeking work rose by 22% between 2001 and 2010,

and even though unemployment in the group nearly doubled — to 2.4% by 2010 — it was still well below the 2010 national average of 8.2%, according to a report from the US National Science Foundation.

Analysts and policy-makers believe that the system is strained to breaking point, and cannot accommodate the existing workforce, let alone sustain continued growth. In an opinion article published in April, former US National Academy of Sciences president Bruce Alberts, founding chair of Harvard Medical School's Department of Systems Biology Marc Kirschner, former president of Princeton University Shirley Tilghman and former US National Institutes of

Health (NIH) director Harold Varmus argue that particularly in the biomedical sciences and in academia, rapid growth has created an unsustainable system (B. Alberts *et al. Proc. Natl Acad. Sci. USA* **111**, 5773–5777; 2014).

The low rate of unemployment does reflect the demand by employers for those with a PhD, but the increasing number of PhDs has seen competition stiffen, even in disciplines traditionally free from such problems.

Tilghman says that she is “frustrated” by the continuing growth in PhD production and notes that she had warned of a pending oversupply back in 1998 in the National Research Council report *Trends in the Early Careers* ►

► *of Life Scientists.* The opinion article, she says, was prompted by the aftermath of the US budget sequester — across-the-board funding cuts that have further strained research budgets and tightened the number of available research posts. The authors call for US universities to gradually cut back the number of students they enrol in biomedical PhD courses and to hire staff scientists, instead of postdocs and graduate students, for day-to-day lab work. More staff-scientist positions would benefit people with PhDs who want to do bench science, and would help principal investigators by providing lab-management assistance (see *Nature* **510**, 433–434; 2014).

Current NIH director Francis Collins says that the agency has begun to address some of these issues, by giving postdocs a 7% pay rise in 2014 and creating a new award for young cancer researchers, for instance. More such awards are in the works, he says.

However, Collins is not convinced that there is an oversupply of life science PhDs. He is also reluctant to alter the way graduate students and postdocs are funded in an attempt to control the PhD pipeline. Shifting them to different types of grant “would be an enormous administrative challenge and would have serious consequences, such as excluding foreign trainees in our workforce”, he says.

Collins agrees with the recommendation that universities tell trainees about a range of career options, not just academic research. Last year, the NIH launched the ‘Broadening Experiences in Scientific Training’ programme to help universities do just that. Information — about both job possibilities and career outcomes — helps trainees to make informed decisions about which path to take, says Collins.

Toby Smith, vice-president of policy for the American Association of Universities in Washington DC, agrees. He says that deans in the association’s membership have begun to ask PhD graduates about what jobs they end up getting. The association is holding a workshop on graduate education in November and it encourages universities to gather data on their alumni.

Keith Micoli, chairman of the US National Postdoctoral Association in Washington DC, agrees that the existing academic-research set-up keeps postdocs dependent on their lab heads for increasingly longer periods. Controlling the supply of young scientists could help, he says, although he worries about unintended consequences. “My fear,” he says, “is that we would just weaken the US science pipeline.” ■

Paul Smaglik is assistant editor of *Nature Careers*.

COLUMN

A wake-up call

Graduate students must educate themselves and others about academia’s dim job prospects, says **Jessica Polka**.



Many graduate students in the biomedical sciences dream of becoming academic researchers. Yet if current employment trends continue, their odds of landing a tenure-track post in the United States are less than 10%, according to data from the US National Institutes of Health (NIH) and the US National Science Foundation. In sum: faculty research positions in this discipline have become an ‘alternative’ career option.

Many biomedical graduate students and their advisers remain woefully ignorant of this development, even though information on the changing workforce is freely available. To consolidate these numbers into a simple picture, I assembled key statistics on the outlook for biomedical PhD holders into an infographic that appeared in April in *The ACSB Post*, the blog for the American Society for Cell Biology (see go.nature.com/vh1ewm). The fact that the trends surprised some of my colleagues suggests that the information has not been sinking in.

Many young scientists, myself included, come to realize the severity of the academic-workforce supply issue only late in their training. This innocence is an unfortunate but natural consequence of the environment we have created.

When a student is applying for a biomedical PhD, she or he is unlikely to learn from programme administrators about the dim prospects for an academic career. Data on trainee outcomes are not readily available, and administrators are not currently required to collect them. The statistics that do exist are sobering, and most university interviewers do not see them as selling points.

Even if provided, the data may at first seem irrelevant to the students. The average trainee takes 7 years to complete a PhD and 4 years to

do a postdoc, according to NIH data, and the volatility of the funding climate makes it impossible to predict the job market 11 (or more) years in advance. The historical context, however, presents a clear picture of an unsustainable system in which the halcyon days are long over.

Also a possibility is that many of our heads are in the sand because we have stuck them there. The numbers sound bad, but we persuade ourselves that they do not apply to us. In a perfect world, we would not have to worry about job security; in the real one, our ignorance seems self-protective. But it isn’t.

It behooves us to not only educate ourselves about the dysfunctions of the system, but also participate in discussions of what to do about it. This is especially true for those of us hell-bent on a faculty position, because at least some of us will be shaping the culture of science. Conversations about creating a sustainable research enterprise are increasingly arising in policy groups, prominent journals, national meetings and online. But even though they are major stakeholders, trainees are conspicuously absent. To make our voices heard, we need to engage in more discussions within scientific societies, student and postdoctoral organizations and online.

Trainees who are vocal about sustainability issues will also help to raise awareness among their peers and the people they mentor, and set a standard of awareness that — we would hope — will one day form a part of US academic culture.

Finally, young scientists must fundamentally rethink their approach to graduate training and postdoctoral appointments. We must see this stage as the only chance many of us will ever have to do biomedical research, rather than as preparation for a well-defined career. But while we take advantage of this unique opportunity, we should also prepare for the future. We must actively explore all professional possibilities; simultaneously, we should advocate for strengthening the future of US science by making it a more attractive career choice. The catalyst for both these changes will be clear dissemination of information on training outcomes to students — and the sooner we start, the better. ■

Jessica Polka is a postdoctoral researcher in systems biology at Harvard Medical School in Boston, Massachusetts.

GORGOLIN/SHUTTERSTOCK

BENJY'S BIRTHDAY

Time to be creative.

BY JOHN GRANT

Lucy loved her husband Hiram but, when he told her at the supper table what their son Benjy was demanding for his 13th birthday, she almost choked on her burger.

"He wants a *what*?" she sputtered.

Nervously, he told her again. "A universe."

"But they're *expensive*!"

"He says all the other kids have them."

"But *their* parents earn the big bucks. Me, I'm strictly middle-management."

"Mark Pollock got one for his tenth," Hiram pointed out. "Benjy's not being unreasonable. Penelope Saunders —"

"That little toad!"

"Penelope Saunders has one. Dwayne Simmons. Lottie McPhee. Denny Barden ... well, Denny Barden's only got a galactic cluster, but Denny's mum's been out of work these past two years. And —"

"Even so."

Hiram shrugged. "Benjy's going to be 13," he said. "This is maybe the last birthday he'll have as a little boy. By next year he could be interested in girls instead."

They shuddered in unison.

"We'll both have to scrimp and save," she said doubtfully.

"We can do it. For our little champion."

Three nights later, when Lucy got home from work, she could tell by the look of suppressed triumph on her husband's face that Hiram had made the purchase. It seemed to take forever for Benjy to get sleepy that evening, and even longer before he could finally be persuaded to drag himself off to bed.

"Let's see it," whispered Lucy when at last they were certain the boy was out for the count.

Hiram led her into the bedroom and opened the wardrobe door.

"Is *that* it?" she said, looking at the box.

"An AC49 universe, just as our son specified," said Hiram proudly. "I was able to get a big discount by paying cash."

"It doesn't look like much, does it?"

The box was maybe a quarter of the size she'd expected. It sat on the floor of the wardrobe looking oddly ... tame, despite the garish pictures on its sides of exploding galaxies and the like.

She bent and picked it up.

"It's so small," she said.

"Apparently it gets a whole lot bigger once you've assembled all the pieces and set it in

motion," Hiram assured her. "That's what the salesman told me. See?" He pointed a stubby finger at the lettering on the box top. "It comes with a built-in inflation kit."

"Well, if you're sure it's the right thing ..."

Lucy bent and put the brightly coloured package back where it had been. Even with the hefty discount, it represented a month's salary. Still, it was a once-in-a-lifetime investment, and worth every cent if it made her little boy happy.

Nevertheless, she slept fitfully that night, waking several times to fret about money.

Luckily Benjy's birthday fell at the weekend that year, so Lucy didn't have to rush off to work through the city's grey dawn. She and Hiram set the alarm for six o'clock and tiptoed around the house getting everything ready for when Benjy awoke. Aside from his main present, there were gifts from his grandparents to be opened, plus a cheque from his aunt that, if past years were anything to go by, would prove to be embarrassingly small.

And then there was the cake that Hiram had baked the day before, when Benjy was out at school. He'd made it in the form of a bright-red racing car, which was probably a mistake but it was too late to do anything about it now.

By the time Benjy eventually straggled downstairs at quarter past eight, his hair sticking out in all directions like a squashed beetle, the living room had been transformed into a veritable Aladdin's cave.

Or so Lucy thought, anyway.

When Benjy saw the box containing his AC49 universe, sitting in pride of place in the middle of the carpet, the sleep vanished from his eyes.

"Wow!"

Lucy and Hiram beamed proudly.

Sitting on the sofa, Benjy turned the box over and over in his hands. He showed no interest at all in any of his other presents. At that moment Lucy didn't care that the next

few months were going to be pretty austere. All that mattered was the awe on her boy's face.

"Assembles in minutes," Benjy said reverentially, reading aloud. "Contains small parts unsuitable for young children. No batteries required."

He slit the tape on the lid with his thumb-nail.

Hours later, the house was silent except for a sort of thrumming sound from the attic, which had been cleared to make space for Benjy's universe. His doting parents had promised him that they wouldn't go up there except in an emergency — that from now on it was his territory, his and his alone.

Lucy and Hiram hadn't quite fallen asleep.

"I didn't expect it to be quite so *noisy* when you were getting it

going," he murmured drowsily.

"It was a hell of a bang, there at the start," she agreed.

"Or for him to take so long setting it up."

"The AC49 comes with the option to fine-tune the physical constants," said Lucy. "Benjy decided he wanted to optimize it properly. Apparently Buster Padawi didn't bother and ended up with something completely chaotic. Our boy wants his universe to support organic life and all. We spent half the morning tinkering with the velocity of light."

Hiram was silent long enough for her to start wondering if he'd drifted off.

"Organic life, huh?" he said, startling her.

"That's the plan."

"Think he'll treat his creations kindly?"

"I very much doubt it," she said, rolling over and squirming herself into a more comfortable position. "Who cares?" ■

John Grant is the author of more than 60 books, both fiction and nonfiction — the latter including such works as *Discarded Science*, *Denying Science*, *The Encyclopedia of Fantasy* (with John Clute) and *A Comprehensive Encyclopedia of Film Noir*; a new collection of his short stories, to be called *Tell No Lies*, is scheduled for late 2014 publication by *Alchemy Press*. He has won the *Hugo* (twice), the *World Fantasy Award* and a bunch of other awards.



JACEY

➔ **NATURE.COM**
Follow Futures:
@NatureFutures
f go.nature.com/mtoodm